On the Ultrametricity in Spin Glass Theory

On EB's Way

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde

(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Nicola Kistler

von

Reichenburg SZ

Promotionskomitee

Prof. Dr. Erwin Bolthausen (Vorsitz)

Prof. Dr. Andrew Barbour

Prof. Dr. Anton Bovier

Zürich 2007

Zusammenfassung

Die Behauptung, dass das Spin-Glass-Verhalten mit hierarchischer Organisation des Konfigurationsraums verbunden sei, die sogenannte Ultrametrizität-Vermutung, ist nur eine jener verblüffenden Vorhersagen der Parisi-Theorie, die trotz Jahrzehnten intensiver Forschung mathematisch noch nicht geklärt sind.

In dieser Doktorarbeit werden Modelle für Spin-Gläser eingeführt, bei welchen die Ultrametrizität mathematisch rigoros untersucht werden kann.

Im ersten Kapitel befassen wir uns mit einer Verallgemeinerung des sogenannten *Generalized Random Energy Model*. Dieses wurde vor zwei Jahrzehnten von Bernard Derrida eingeführt und hat seither eine entscheidende Rolle bei der mathematischen Aufklärung der rätselhaften Parisi-Theorie gespielt. Das GREM ist aber ein Spin-Glass auf Bäumen (von vornherein hierarchisch organisiert) und kann demzufolge in Bezug auf die Ultrametrizität nur von geringem Nutzen sein. In unseren Verallgemeinerungen sind beliebige Graphen-Strukturen zugelassen.

Nebst seiner pädagogischen Bedeutung ist das GREM in der Parisi-Theorie von zentralem Belang, da es als *universeller Attraktor* in Erscheinung tritt. Das sollte zumindest für ungeordnete Systeme der Klasse von Sherrington und Kirkpatrick der Fall sein, für welche eine hierarchische Brechung der Symmetrie *à la Parisi* zu einer Freien Energie führt, die einer Spin-Störung einer GREM-Struktur entspricht. Trotz dieses ansprechenden Bildes ist es bis heute noch unklar, wie und in welchem Ausmass solche Strukturen im thermodynamischen Limes entstehen.

Gewisse Aspekte dieses Phänonems können mit Hilfe eines im zweiten Kapitel eingeführten Modells erklärt werden, welches aus einem REM, dem einfachsten Modell von Derrida, und einem sogenannten *Cavity Field* besteht. Dank zugrundeliegenden, universellen Mechanismen, übertragen sich unsere Resultate auf grundsätzlich jedes Spin-Glass-Modell unendlicher Reichweite, für welches REM-artige Strukturen vorhanden sind.

Um Modelle mit komplexeren, GREM-artigen Abhängigkeitsstrukturen untersuchen zu können, versuchen wir im dritten und letzten Kapitel die Reichweite unseres Leitbildes zu erweitern. Jedoch nur teilweise mit Erfolg: obwohl es uns gelungen ist, das Problem auf der Stufe der Freien Energie zu lösen, konnten wir das Verhalten des Gibbs-Masses nicht enträtseln.

Introduction

The claim that Spin Glass behavior comes along with hierarchical organization of the configuration space is often referred to as the *ultrametricity conjecture*. It is one of those striking predictions of the Parisi Theory for Spin Glasses which remains, despite decades of intensive research, mathematically very poorly understood.

In this thesis we introduce some mean field models for Spin Glasses where the issue of ultrametricity can be addressed in complete mathematical rigor.

In the first chapter we consider natural extensions of the Generalized Random Energy Model, GREM for short. The latter had been introduced more than 20 years ago by Bernard Derrida and has played ever since a crucial rôle in the attempt to gain some understanding in the puzzling Parisi Theory. Being a spin glass on a tree, the GREM gives little clue pertaining to the *onset* of ultrametricity when the thermodynamical limit is taken. In our generalizations we allow for virtually any underlying graph structure.

Beside its pedagogical relevance, the GREM has proved of central importance in the Parisi Theory, coming forward as a "universal attractor" for disordered systems. This *seems* to be the case for instance in the class of Sherrington-Kirkpatrick models, where hierarchical Replica Symmetry Breaking leads to a free energy given by a one spin perturbation of some properly chosen GREM, after the thermodynamical limit has been taken. Despite this appealing picture, it remains a mystery why and in which way the scenario sets in with the thermodynamical limit.

We provide some modest clarification of this issue in the second chapter by solving a system obtained by a cavity field perturbation of the simplest Derrida's model, the REM. Nicely enough, our results extend in virtue of some underlying abstract Theorems to any spin glass with REM-type structure, shedding some light on the alleged universality of the Parisi Theory.

In order to cover systems with more intricate dependencies, in particular of GREM-type, we push the abstract approach a bit further in the third chapter. Quite an endeavour: though we settle the problem at the level of the free energy (providing an unexpected interpretation for the Order Parameter of the Parisi Theory), we were not able to settle the issue at the level of the Gibbs measure.

Danksagung

Ich möchte mich an dieser Stelle bei allen bedanken, die mich während meiner Zeit als Doktorand unterstützt haben.

Ein grosser Dank geht an meinen Betreuer, Prof. Erwin Bolthausen, für die Möglichkeit, Ihn ein paar Schritte auf dem Weg zur Ultrametrizität begleiten zu dürfen. Ebenso möchte ich Prof. Anton Bovier dafür danken, dass er diese Dissertation begutachtet hat.

Meine Familie und Freunde waren immer bedingungslos an meiner Seite. An Sie alle das grösste Dankeschön.

Nicola Kistler, Dezember 2006

Contents

1. On a non hierarchical version of the Generalized Random Energy Model 1
   - 1.1. Introduction of the model and main results 1
   - 1.2. Free energy of the non hierarchical GREM, Proofs 20
   - 1.3. Gibbs measure of the non hierarchical GREM, Proofs 26
2. On a cavity field perturbation of the Random Energy Model 45
   - 2.1. Introduction and outline 45
   - 2.2. Mean field models of REM-type 45
   - 2.3. The REM+Cavity 47
   - 2.4. Proofs of the main results 50
3. On a cavity field Perturbation of the GREM 65
   - 3.1. Introduction and outline 65
   - 3.2. Mean field models of GREM-type 65
   - 3.3. An application, the GREM+Cavity 67
   - 3.4. Proofs of the main results 69

Bibliography 81

Curriculum Vitae 83

1. On a non hierarchical version of the Generalized Random Energy Model

1.1. Introduction of the model and main results

The Generalized Random Energy Model (GREM for short), introduced by B. Derrida in [**11**] plays in important rôle in spin glass theory. Originally invented as a simple model which exhibits replica symmetry breaking at various levels, it has become clear that more interesting models, like the celebrated one of Sherrington-Kirkpatrick, exhibit GREM-like behavior in the large N limit. Despite the spectacular recent progress in understanding the SK-model (see [**13**], [**14**], [**20**]), many issues have not been clarified at all, the most prominent one being the so called *ultrametricity.*[1] The GREM is of limited use to investigate this because it is hierarchically organized from the start. This favorable situation allows for a complete solution, fully confirming the so called Parisi theory (we refer the reader to the detailed study [**6**], where it is also pointed out that, interestingly, the emerging ultrametricity of the Gibbs measure does not necessarily coincide with the starting hierarchical organization). Yet, from the considerations on the GREM, one gets little clue on why many systems should be ultrametric in the limit [2].

We present here a simple, and as we think natural, generalization of the GREM which has *no* built in ultrametric structure. Our main result is that the limiting free energy always coincides with that of a suitably constructed GREM. This is not the case at the level of the Gibbs measure; in fact, we prove that genuine ultrametricity only holds provided some irreducibility conditions on the hamiltonian are met, in which case we prove the "full Parisi Picture", thereby establishing the law of the Gibbs measure in low temperature, the law of the overlap, and the link to the Bolthausen-Sznitman coalescent introduced in [**4**].

We begin fixing a number $n \in \mathbb{N}$, and consider the set $I = \{1, ..., n\}$, as well as a collection of positive real numbers $\{a_J\}_{J\subset I}$ such that

$$\sum_{J\subset I} a_J = 1.$$

[1] A metric d is called an *ultrametric,* if the strengthened triangle condition holds: $d(x,z) \leq \max(d(x,y), d(y,z))$. Equivalently, two balls are either disjoint or one is contained in the other.

[2] In Talagrand's recent proof of the Parisi formula, ultrametricity plays no apparent rôle, and it seems to be quite delicate to prove ultrametricity by Talagrand's method. This is quite curious as on the other hand, ultrametricity plays a *crucial* rôle in the physicists non-rigorous derivation of the free energy, be that using the replica trick or the cavity method.

For convenience, we put $a_\emptyset \overset{\text{def}}{=} 0$. The relevant subset of I will be only the ones with positive a-value. For $A \subset I$ we set

$$\mathcal{P}_A \overset{\text{def}}{=} \{J \subset A : a_J > 0\},\ \mathcal{P} \overset{\text{def}}{=} \mathcal{P}_I.$$

For $n \in \mathbb{N}$, we set $\Sigma_N \overset{\text{def}}{=} \left\{1, \ldots, 2^N\right\}$. We also fix positive real numbers γ_i, $i \in I$, satisfying

$$\sum_{i=1}^{n} \gamma_i = 1.$$

The collection $(a_J, \gamma_i; J \in \mathcal{P}, i \in I)$ will be referred to as *underlying parameters.* We write $\Sigma_N^i \overset{\text{def}}{=} \Sigma_{\gamma_i N}$, where for notational convenience we assume that $2^{\gamma_i N}$ is an integer. For $N \in \mathbb{N}$, we will label the "spin configurations" σ as:

$$\sigma = (\sigma_1, ..., \sigma_n),\ \sigma_i \in \Sigma_N^i\ ,$$

i.e. we identify Σ_N with $\Sigma_N^1 \times \cdots \times \Sigma_N^n$. For $J \subset I$, $J = \{j_1, \ldots, j_k\}$, $j_1 < j_2 < \ldots < j_k$, we write $\Sigma_{N,J} \overset{\text{def}}{=} \prod_{s=1}^k \Sigma_N^{j_s}$, and for $\sigma \in \Sigma_N$, we write σ_J for the projected configuration $(\sigma_j)_{j \in J} \in \Sigma_{N,J}$. Our spin glass Hamiltonian is defined as

$$X_\sigma = \sum_{J \in \mathcal{P}} X_{\sigma_J}^J \tag{1.1}$$

where $X_{\sigma_J}^J$, $J \in \mathcal{P}$, $\sigma_J \in \Sigma_{N,J}$ are independent centered Gaussian random variables with variance $a_J N$. The X_σ are then Gaussian random variables with variance N (Gaussian always means "centered Gaussian" through this note), but they are correlated. $\mathbb{E}$ will denote expectation with respect to these random variables. A special case is when $\mathcal{P} = \{I\}$, i.e. when only $a_I \neq 0$, in which case it has to be one. Then the X_σ are independent, i.e. one considers simply a set of 2^N independent Gaussian random variables with variance N. This is the standard Random Energy Model.

The Generalized Random Energy Model is a special case, too: It corresponds to the situation where the sets in $\mathcal{P}$ are "nested", meaning that $\mathcal{P}$ consists of an increasing sequence of subsets. Without loss of generality we may assume that in this case

$$\mathcal{P} = \{J_m : 1 \leq m \leq k\},\ J_m \overset{\text{def}}{=} \{1, \ldots, n_m\}, \tag{1.2}$$

where $1 \leq n_1 < n_2 < \ldots < n_k \leq n$. In the GREM case, the natural metric on Σ_N coming from the covariance structure:

$$d(\sigma, \sigma') \overset{\text{def}}{=} \sqrt{\mathbb{E}\left((X_\sigma - X_{\sigma'})^2\right)}$$

is an *ultrametric.* In the more general case (1.1) considered here, this metric is *not* an ultrametric.

To see this, take $n=3$, $\mathcal{P}=\{\{1,2\},\{1,3\},\{2,3\}\}$, i.e. where

$$X_\sigma = X^{\{1,2\}}_{\sigma_1,\sigma_2} + X^{\{1,3\}}_{\sigma_1,\sigma_3} + X^{\{2,3\}}_{\sigma_2,\sigma_3} \tag{1.3}$$

with $a_J = 1/3$ for $J \in \mathcal{P}$. Then for $a,b,b',c,c' \in \Sigma_{N/3}$, $b \neq b'$, $c \neq c'$, one has

$$d\left((a,b,c),(a,b,c')\right) = d\left((a,b,c'),(a,b',c')\right) = \sqrt{2N/3},$$

whereas

$$d\left((a,b,c),(a,b',c')\right) = \sqrt{N},$$

contradicting ultrametricity.

1.1.1. The free energy of the non hierarchical GREM. Any of our models can be "coarse-grained" in many ways into a GREM. For that consider strictly increasing sequences of subsets of $I : \emptyset = A_0 \subset A_1 \subset \ldots \subset A_K = I$. We do not assume that the A_i are in $\mathcal{P}$. We call such a sequence a **chain** $\mathbf{T} = (A_0, A_1, \ldots, A_K)$. We attach weights $\hat{a}_{A_j}$ to these sets by putting

$$\hat{a}_{A_j} \stackrel{\text{def}}{=} \sum_{B \in \mathcal{P}_{A_j} \setminus \mathcal{P}_{A_{j-1}}} a_B, \tag{1.4}$$

Evidently $\sum_{j=1}^K \hat{a}_{A_j} = 1$, and if we assign random variables $X_\sigma(\mathbf{T})$, according to (1.1), we arrive after an irrelevant renumbering of I at a GREM of the form (1.2). In particular, the corresponding metric d is an ultrametric.

We write $\mathrm{tr}\,(\cdot)$ for averaging over Σ_N (i.e. the coin-tossing expectation if we identify Σ_N with $\{H,T\}^N$).

For a function $x : \Sigma_N \to \mathbb{R}$, set

$$Z_N(\beta, x) \stackrel{\text{def}}{=} \mathrm{tr} \exp[\beta x], \; F_N(\beta, x) \stackrel{\text{def}}{=} \frac{1}{N} \log(Z_N(\beta, x)),$$

and define the usual finite N partition function, and free energy by

$$Z_N(\beta) \stackrel{\text{def}}{=} Z_N(\beta, X), \; F_N(\beta) \stackrel{\text{def}}{=} F_N(\beta, X), \; f_N(\beta) \stackrel{\text{def}}{=} \mathbb{E}\left(F_N(\beta, X)\right),$$

where X is interpreted as random function $\Sigma_N \to \mathbb{R}$.

For any chain $\mathbf{T}$, we attach to our model a GREM $(X_\sigma(\mathbf{T}))_{\sigma \in \Sigma_N}$, as explained above, and then

$$f_N(\mathbf{T}, \beta) \stackrel{\text{def}}{=} \mathbb{E}\left(F_N(\beta, X(\mathbf{T}))\right),$$

$$f(\mathbf{T}, \beta) \stackrel{\text{def}}{=} \lim_{N \to \infty} f_N(\mathbf{T}, \beta).$$

For a GREM, the limiting free energy is known to exist, and can be expressed explicitly, but in a somewhat complicated way (see [**11**], [**8**]). Our main result is that our generalization of the GREM does not lead to anything new in $N \to \infty$ limit, shedding hopefully some modest light on the "universality" of ultrametricity.

Theorem 1.1.

$$f(\beta) \stackrel{\text{def}}{=} \lim_{N\to\infty} f_N(\beta) \tag{1.5}$$

exists, and is also the almost sure limit of $F_N(\beta)$.

$f(\beta)$ *is the free energy of a GREM. More precisely, there exists a chain* $\mathbf{T}$ *such that*

$$f(\beta) = f(\mathbf{T}, \beta), \ \ \beta \geq 0. \tag{1.6}$$

$f(\mathbf{T}, \beta)$ *is minimal in the sense that*

$$f(\beta) = \min_{\mathbf{S}} f(\mathbf{S}, \beta), \tag{1.7}$$

the minimum being taken over all chains $\mathbf{S}$.

The fact the free energy is *self-averaging,* meaning that $f(\beta)$ (if the limit exists) is also the almost sure limit of the F_N is a simple consequence of the Gaussian concentration inequality. We write F_N as a function of the standardized variables $X^J_{\sigma_J}/\sqrt{a_J N}$. As

$$\left|\log \sum\nolimits_i \mathrm{e}^{a_i} - \log \sum\nolimits_i \mathrm{e}^{a_i'}\right| \leq \max_i |a_i - a_i'|, \ \ a_i, a_i' \in \mathbb{R},$$

we get that $F_N(\beta)$, regarded as a function of the collection $\left(X^J_{\sigma_J}/\sqrt{a_J N}\right)$ is Lipschitz continuous with Lipshitz constant $\beta/\sqrt{N}$. By the usual concentration of measure estimates for Gaussian distributions (see e.g. Proposition 2.18 of [**15**]), we have

$$\mathbb{P}[|F_N(\beta) - \mathbb{E}F_N(\beta)| > \epsilon] \leq 2\exp\left[-\frac{\epsilon^2}{2\beta^2}N\right] \tag{1.8}$$

Using the Borel-Cantelli lemma, one sees that if $\lim_{N\to\infty} f_N(\beta)$ exists, then the $F_N(\beta)$ converge almost surely to this limit, too, and if $\lim_{N\to\infty} F_N(\beta)$ exists almost surely, then the limit is non-random and equals $\lim_{N\to\infty} f_N(\beta)$.

The existence of the limit is established through a quite standard application of the *Second Moment Method*, akin to that originally exploited by Derrida in his seminal paper [**11**]; this allows to express the limiting free energy in terms of a variational problem, which we then solve inductively.

For the reader's convenience, we briefly describe the mechanism which lies behind Theorem 1.1 for the Hamiltonian (1.3) but we allow for general (positive) variances a_{12}, a_{13}, a_{23}, and general γ_i. It is best to count the number of configurations σ which reach a certain energy level λN. It is evident that only an exponentially small portion of the total number 2^N of configurations achieve this, roughly formulated (we will be more precise later)

$$\#\left\{\sigma : X_\sigma \simeq \lambda N\right\} \simeq 2^N \mathrm{e}^{-\rho(\lambda)N}, \ \ \rho(\lambda) > 0.$$

The free energy is obtained by the Legendre transform of ρ. In order to determine $\rho(\lambda)$ we count individually for each of the three parts in (1.3) how many configurations reach respective levels

$$\hat{\rho}(\lambda_1,\lambda_2,\lambda_3) \simeq -\frac{1}{N}\log\#\left\{\sigma : X^{\{1,2\}}_{\sigma_1,\sigma_2} \simeq \lambda_1 N,\ X^{\{1,3\}}_{\sigma_1,\sigma_3} \simeq \lambda_2 N,\ X^{\{2,3\}}_{\sigma_2,\sigma_3} \simeq \lambda_3 N\right\} + \log 2 \tag{1.9}$$

with $\lambda_1+\lambda_2+\lambda_3=\lambda$. Evidently

$$\rho(\lambda) = \inf_{\lambda_1+\lambda_2+\lambda_3=\lambda} \hat{\rho}(\lambda_1,\lambda_2,\lambda_3). \tag{1.10}$$

It turns out that one can get $\hat{\rho}$ by computing expectations inside the logarithm, provided only some naturally defined restrictions on the λ_i are satisfied. For small λ, it is easily seen that one has an "equipartition" property, and that the optimal $\lambda_1,\lambda_2,\lambda_3$ are proportional to the respective variances, i.e. $\lambda_1 = a_{12}\lambda,\ \lambda_2 = a_{13}\lambda$, $\lambda_3 = a_{23}\lambda$, and from that one obtains

$$\rho(\lambda) = \lambda^2/2, \tag{1.11}$$

which is the same as if the X_σ would be uncorrelated. Increasing λ, we however encounter restrictions from the structure of the Hamiltonian. First of all, λ_1 has to be such that there *are* any σ_1,σ_2 with $X^{\{1,2\}}_{\sigma_1,\sigma_2} \simeq \lambda_1 N = a_{12}\lambda N$. There are $2^{(\gamma_1+\gamma_2)N}$ pairs (σ_1,σ_2), and as the $X^{\{1,2\}}_{\sigma_1,\sigma_2}$ are independent, the restriction is

$$2^{(\gamma_1+\gamma_2)N}\exp\left[-\frac{\lambda^2 a_{12} N}{2}\right] \gtrapprox 1.$$

(We are not considering any log-corrections). This leads to the restriction

$$\lambda \le \sqrt{\frac{2(\gamma_1+\gamma_2)\log 2}{a_{12}}} \tag{1.12}$$

for the validity of (1.11), and there are two similar restrictions coming from $X^{\{1,3\}}$ and $X^{\{2,3\}}$. Even if these three restrictions are satisfied, it can be that there are simply totally not enough triples $(\sigma_1,\sigma_2,\sigma_3)$ left. A *necessary* condition for this is certainly that the expected number of $\#\{\sigma : X_\sigma \simeq \lambda N\}$ is not exponentially decaying, which is simply the condition that $\lambda \le \sqrt{2\log 2}$. The somewhat astonishing fact is that these are the only conditions one has to take into considerations, for the validity of (1.11). Now, there are two cases:

Case 1: $\lambda \le \sqrt{2\log 2}$ implies the other ones, i.e.

$$\min_{1\le i<j\le 3} \frac{\gamma_i+\gamma_j}{a_{ij}} \ge 1. \tag{1.13}$$

In that case, we are simply left with the restriction $\lambda \le \sqrt{2\log 2}$, and the free energy is

$$f(\beta) = \sup_{\lambda\le\sqrt{2\log 2}} \left(\beta\lambda - \lambda^2/2\right),$$

which is the free energy of a REM. In that case the internal structure of the model is irrelevant, at least for the free energy.

Case 2: (1.13) is violated. For definiteness, assume that $(\gamma_1+\gamma_2)/a_{12}$ is the smallest one.

In that case, (1.11) is only correct in the region (1.12). For λ larger, there is no (σ_1,σ_2) with $X^{\{1,2\}}_{\sigma_1,\sigma_2} \simeq a_{12}\lambda N$ (with probability close to 1), the maximum of the $X^{\{1,2\}}_{\sigma_1,\sigma_2}$ being at $m_{12}N$ ($\pm$ log-corrections), where

$$m_{12} \stackrel{\text{def}}{=} \sqrt{2(\gamma_1+\gamma_2)\,a_{12}\log 2}$$

Therefore, one has to restrict in (1.10) to λ's with $\lambda_1 = m_{12}$. The only configurations σ for which $X_\sigma \simeq \lambda N$ have to satisfy

$$X^{\{1,2\}}_{\sigma_1,\sigma_2} \simeq m_{12}N, \tag{1.14}$$

but there are now only subexponentially many (σ_1,σ_2) left which achieve this feat, and the difference to λN has to be made by the field

$$Y_{\sigma_1,\sigma_2}(\sigma_3) \stackrel{\text{def}}{=} X^{\{1,3\}}_{\sigma_1,\sigma_3} + X^{\{2,3\}}_{\sigma_2,\sigma_3},\ 1\le\sigma_3\le 2^{\gamma_3 N},$$

restricting (σ_1,σ_2) to the few which satisfy (1.14). There is an upper limit $\lambda_{\max}$ for λ's such that there are *any* σ_3 with $Y(\sigma_3) \simeq (\lambda - m_{12})N$. $\lambda_{\max}N - m_{12}N$ is simply the maximum of $2^{\gamma_3 N}$ independent Gaussians with variance $(a_{13}+a_{23})N$, i.e.

$$\lambda_{\max} - m_{12} \stackrel{\text{def}}{=} \sqrt{2\gamma_3(a_{13}+a_{23})\log 2}.$$

The situation is similar to the one in the GREM with the only difference that for $(\sigma_1,\sigma_2) \neq (\sigma_1',\sigma_2')$, the fields Y_{σ_1,σ_2} and $Y_{\sigma_1',\sigma_2'}$ are not independent, except when $\sigma_1\neq\sigma_1'$ *and* $\sigma_2\neq\sigma_2'$. It is however fairly evident that among the (σ_1,σ_2) for which $X^{\{1,2\}}_{\sigma_1,\sigma_2}\simeq m_{12}N$ there will be no pairs with such a partial overlap, with probability close to 1, and therefore it is quite natural one can handle the field Y_{σ_1,σ_2} as if it would come from a second level of a two-level GREM. In fact it turns out that in the Case 2, the tree of Theorem 1.1 is $\{\{1,2\},\{1,2,3\}\}$, and we replace our model with the coarse grained one with Hamiltonian $X'_{\alpha_1}+X''_{\alpha_1,\alpha_2}$, where $\alpha_1 = 2^{(\gamma_1+\gamma_2)N}$, $\operatorname{var}(X') = a_{12}N$, $\#\alpha_2 = 2^{\gamma_3 N}$, $\operatorname{var}(X'') = (a_{13}+a_{23})N$.

This way of reasoning works for the general case. There are some issues which might be somewhat surprising. The first is that expressions (1.9) can always be evaluated by computing expectations inside the logarithm, provided one keeps some fairly trivial restrictions on the λ_i. Secondly, it is not entirely evident why (and to which extent) these restrictions finally always lead to tree structure. It is also interesting that the system always chooses from the many GREMs which can be obtained by coarse-grainings the one with minimal free energy; a similar behavior had already been observed for the GREM itself, cfr. [**6**].

The variational problem for the free energy of the non hierarchical GREMs as in Theorem 1.1 can be solved explicitely through a simple procedure, yielding a strictly increasing sequence of subsets $A_0 \stackrel{\text{def}}{=} \emptyset \subset A_1 \subset A_2 \subset ... \subset A_K = I$ and parameters $\beta_0 \stackrel{\text{def}}{=} 0 < \beta_1 < \beta_2 < ... < \beta_K < \infty$ by recursion. Let us first write, for $A \subset I = \{1, \dots, n\}$,

$$\alpha(A) \stackrel{\text{def}}{=} \sum_{J \in \mathcal{P}_A} a_J, \quad \gamma(A) \stackrel{\text{def}}{=} \sum_{i \in A} \gamma_i.$$

For $B \subset A$ let

$$\rho(B, A) \stackrel{\text{def}}{=} \sqrt{2 \log 2 \frac{\gamma(A) - \gamma(B)}{\alpha(A) - \alpha(B)}}, \quad \hat{\rho}(B) \stackrel{\text{def}}{=} \min_{A : A \supset B, A \neq B} \rho(B, A).$$

Assume that $\emptyset \subset A_1 \subset A_2 \subset ... \subset A_k$ and $0 < \beta_1 < \beta_2 < ... < \beta_k$ are constructed such that the following conditions are satisfied:

C1(k) $\beta_j = \hat{\rho}\left(A_{j-1}\right), \ j \leq k$, and

C2(k) For $j \leq k$ and any $A \supset A_{j-1}$ which satisfies $\beta_j = \rho\left(A_{j-1}, A\right)$ one has $A \subset A_j$, i.e. A_j is *maximal* with $\beta_j = \rho\left(A_{j-1}, A_j\right)$.

This "algorithmic construction" stops after K steps, where K is simply given by the relation $A_K = I$. The extremal chain is then given by $\mathbf{T} = \{A_1, \dots, A_K\}$. Moreover, the sequence $\{\beta_1, \dots, \beta_K\}$ corresponds to the inverse of temperatures where phase transitions at the level of the free energy occur:

Proposition 1.2. *With the above notations, the free energy of the non hierarchical GREM is given by*

$$f(\beta) = \beta \sum_{i=1}^{k} \beta_i \left[\alpha(A_i) - \alpha(A_{i-1})\right] - \gamma(A_k) \log 2 + \frac{\beta^2}{2}(1 - \alpha(A_k)) \tag{1.15}$$

for $\beta_k \leq \beta \leq \beta_{k+1}$.

1.1.2. The Gibbs measure of the non hierarchical GREM. To address the issue of the Gibbs measure we need to construct an infrastructure allowing in the end to attach marks to a Point Process with the random mechanism which governs the marks being independent of the Point Process itself. As curious as it may sound, this is one of the astonishing predictions of the Parisi Theory.

Let X be a locally compact space with countable base (lccb for short). We write $\mathcal{M}(X)$ for the set of Radon measures, and $\mathcal{M}_p(X)$ for the subset of pure point measures. We also write $X^{(2)}$ for the set of two-element subsets of X. Clearly,

$X^{(2)}$ is a lccb, too [*we can identify it for instance with* $(X^2 \setminus D)/\sim$ *, where* D *is the diagonal* $\{(x,x) : x \in X\}$ *and* $(x,y) \sim (y,x)$]. We write π for the projection $(X^2 \setminus D) \to X^{(2)}$.

Any Radon measure μ on X induces a Radon measure $\mu^{(2)}$ on $X^{(2)}$ by first taking the product measure $\mu \times \mu$ on X^2, restrict it to the complement of the diagonal, and project it on $X^{(2)}$. We write $\psi : \mathcal{M}(X) \to \mathcal{M}(X)$ for this mapping. The image of a pure point measure is clearly a pure point measure. Also, if K is a compact subset of X, then $\rho_K : \mathcal{M}(X) \to \mathcal{M}(K)$ is given by restricting $\mu \in \mathcal{M}(X)$ to K. This transforms pure point measures in pure point measures, of course. For compact K and $\mu \in \mathcal{M}(K)$, the total mass $|\mu|$ of μ is finite. If $\mu \in \mathcal{M}_p(K)$, this is just the number of points of the point measure μ. *[It is easy to see that* ψ *is continuous in the vague topology. For this, consider a continuous function with bounded support* $f : X^{(2)} \to \mathbb{R}$*. Then* $f \circ \pi$ *has compact support on* $(X^2 \setminus D)$ *and therefore, we can extend it (by* 0*) to a function of compact support on* X^2*, which we still write as* $f \circ \pi$*. Assume* $\mu_n \to \mu$ *vaguely, for* $\mu_n, \mu \in \mathcal{M}(X)$*. Then* $\lim_{n\to\infty} \int f d\mu_N^{(2)} = \lim_{n\to\infty} \int f \circ \pi d\mu_n = \int f d\mu^{(2)}$*.]*

Let F be a finite set (this suffices for our purposes). If Y is a lccb, we define $\mathcal{M}_{mp}(Y \times F)$ to be the subset of $\mathcal{M}_p(Y \times F)$ consisting of measures with the property that its marginal on Y is in $\mathcal{M}_p(Y)$. In other words, the measures in $\mathcal{M}_{mp}(Y \times F)$ are of the form

$$\sum_i \delta_{\{y_i, a_i\}}, \quad y_i \in Y,\ a_i \in F$$

where the y_i are all distinct, and $\{y_i\}$ is locally finite. It is clear that $\mathcal{M}_{mp}(Y \times F)$ is a measurable subset of $\mathcal{M}(Y \times F)$. Weak convergence of probability measures on $\mathcal{M}_{mp}(Y \times F)$ refers to weak convergence of their extensions to $\mathcal{M}(Y \times F)$.
If $K \subset Y$ is a compact subset, then we set $\hat{\rho}_K : \mathcal{M}_{mp}(Y \times F) \to \mathcal{M}_{mp}(K \times F)$ by taking the restriction. It is clear that any probability measure P on $\mathcal{M}_{mp}(Y \times F)$ is uniquely determined by the family $P\hat{\pi}_K^{-1}$, K compact in Y. Furthermore, for any consistent family of such probability measures P_K on $\mathcal{M}_{mp}(K \times F)$, $K \subset\subset Y$, there is a unique probability measure P on $\mathcal{M}_{mp}(Y \times F)$ with $P\hat{\pi}_K^{-1} = P_K$. Consistency means that for $K \subset K'$ one has $P'_K \hat{\pi}_{K',K}^{-1} = P_K$, where $\pi_{K',K} : \mathcal{M}_{mp}(K' \times F) \to \mathcal{M}_{mp}(K \times F)$. This follows easily from Kolmogoroff's Theorem. It suffices to have the P_K consistently defined for a sequence of compacta (K_n) with $K_n \uparrow Y$.

Let $\mathbb{N}^{(2)} \stackrel{\text{def}}{=} \{(i,j) : \ i,j \in \mathbb{N}, i < i\}$. We consider probability measures Q on $F^{\mathbb{N}^{(2)}}$ which have the property that they are invariant under finite permutations: a permutation $\mathfrak{s} : \mathbb{N} \to \mathbb{N}$ which leaves all the number except finitely many fixed induces a mapping $\phi_{\mathfrak{s}} : F^{\mathbb{N}^{(2)}} \to F^{\mathbb{N}^{(2)}}$ in a natural way. We call Q invariant if it is invariant under all such $\phi_{\mathfrak{s}}$.

Given a sequence of distinct points $\mathbf{x} = (x_1, \ldots, x_N)$ in some compact set K, and $\mathbf{f} = (f_{ij}, 1 \leq i < j \leq N)$, $f_{ij} \in F$, we put

$$L(\mathbf{x}, \mathbf{f}) \stackrel{\text{def}}{=} \sum_{i<j} \delta_{x_i, x_j, f_{ij}} \in \mathcal{M}_{mp}(K^{(2)} \times F).$$

For fixed $\mathbf{x}$, this defines a mapping $L(\mathbf{x}, \cdot) : F^{\hat{N}} \to \mathcal{M}_{mp}(K^{(2)} \times F)$, where $\hat{N} \stackrel{\text{def}}{=} \{(i,j) : 1 \leq i < j \leq N\}$. Given an invariant Q on $F^{\mathbb{N}^{(2)}}$, $N \in \mathbb{N}$, we write Q_N for its restriction on $F^{\hat{N}}$. The $Q_N L(\mathbf{x}, \cdot)^{-1}$ is a probability measure on $\mathcal{M}_{mp}(K^{(2)} \times F)$, depending still on N and $\mathbf{x}$. We denote it by $\Pi(N, \mathbf{x}; \cdot)$. By the invariance property of Q, it only depends on the set $\{x_1, \ldots, x_N\}$ (or on $\sum \delta_{x_i}$). Therefore, for fixed N, $\Pi(N, \cdot; \cdot)$ is a Markov Kernel from $\mathcal{M}_{p,N}(K) \stackrel{\text{def}}{=} \{\mu \in \mathcal{M}_{p,N}(K) : |\mu| = N\}$ to $\mathcal{M}_{mp}(K^{(2)} \times F)$.

With X lccb, and P a probability on $\mathcal{M}_p(X)$, we choose compacts (K_n) with $K_n \uparrow X$. We also write $P_n \stackrel{\text{def}}{=} P\rho_{K_n}^{-1}$ on $\mathcal{M}_p(K_n)$. Then we define $\hat{P}_n$ on $\mathcal{M}_{mp}(K_n^{(2)} \times F)$ by

$$\hat{P}_n \stackrel{\text{def}}{=} \int P_n(d\mu) \Pi(|\mu|, \mu; \cdot).$$

This satisfies the above required consistency property, and therefore gives rise to a probability measure on $\mathcal{M}_{mp}(X^{(2)} \times F)$, which evidently does not depend on the sequence (K_n) chosen, and is denoted by $P \sqcap Q$.

We specify F further, by choosing it to be the set, 2^I, of subsets of the finite set $I = \{1, \ldots, n\}$. We write $\mathcal{G}_{\beta,N}(\sigma) \stackrel{\text{def}}{=} Z_N^{-1}(\beta) \exp\left[\beta X_\sigma\right]$ for the quenched Gibbs measure; with $\mathbf{T} = (A_0, A_1, \ldots, A_K)$ the extremal chain from Theorem 1.1, and $m = 1, \ldots, K-1$, we denote by $\mathbf{T}^{(m)} \stackrel{\text{def}}{=} (A_0, \ldots, A_{m-1}, A_m)$ the chain restricted to the first "m-levels". For $\sigma, \sigma' \in \Sigma_N$ we define the overlap with values in 2^I to be the subset where they agree: $q(\sigma, \sigma') \stackrel{\text{def}}{=} \{i \in I : \sigma_i = \sigma'_i\}$. A fixed realization of the Hamiltonian induces an element of $\mathcal{M}_{mp}\Big((\mathbb{R}^+)^{(2)} \times 2^I\Big)$ by setting

$$\sum_{\sigma,\sigma'} \delta_{\{\mathcal{G}_{N,\beta}(\sigma), \mathcal{G}_{N,\beta}(\sigma'); q(\sigma,\sigma')\}}.$$

We denote by $\Xi_{N,\beta}$ its distribution under $\mathbb{P}$. under $\Xi_{N,\beta}^{(m)}$ we understand the law of the element of $\mathcal{M}_{mp}\Big((\mathbb{R}^+)^{(2)} \times 2^{A_m}\Big)$ induced by the m^{th}-marginal of the Gibbs measure, the latter being the collection of points

$$\mathcal{G}^{(m)}_{\beta,N}(\tau) \stackrel{\text{def}}{=} \sum_{\sigma\in\Sigma_N:\sigma_{A_m}=\tau} \mathcal{G}_{\beta,N}(\sigma), \quad \tau \in \Sigma_{N,A_m}.$$

We now recall the basic construction of the coalescent on $\mathbb{N}$ introduced in [**5**]. This is a continuous time Markov process $(\psi_t, t \geq 0)$ taking values in the compact set of partitions on $\mathbb{N}$. We call a partition $\mathcal{C}$ finer than $\mathcal{D}$, in notation $\mathcal{C} \succ \mathcal{D}$ provided that the sets of $\mathcal{D}$ are unions of the sets of $\mathcal{C}$. The process $(\psi_t, t \geq 0)$ has the following properties

- If $t \geq s$ then $\psi_s \succ \psi_t$.
- The law of $(\psi_t, t \geq 0)$ is invariant under permutations.
- $\psi_0 = 2^{\mathbb{N}}$.

We denote the equivalence relation associated with ψ_t by $\sim_t$. Given this coalescent, a sequence $\mathbf{t} = (t_0, \ldots, t_K)$ of *times* $t_0 = 0 < t_1 < t_2 \cdots < t_{K-1} < t_K = \infty$, and a chain $\mathbf{T}$ as above, we attach to each pair $i < j$ of natural numbers randomly the $A_{K-m}, 1 \leq m \leq K$ (and only these) where $m \stackrel{\text{def}}{=} \min\{l : i \sim_{t_l} j\}$. This defines a law $Q_{\mathbf{T},\mathbf{t}}$ on $(2^I)^{\mathbb{N}^{(2)}}$.

Before stating our main result, we stress once again that it is not possible to extend the one-to-one correspondence between the non hierarchical models and the GREMs to the finer properties of the systems, such as that of the Gibbs measure, at least not in full generality; in fact, it turns out that the limiting systems are genuinely ultrametric only provided some *irreducibility conditions* on the hamiltonian are met, motivating the following definition:

Definition 1.3. *We say that an hamiltonian is irreducible if the following holds:*

$\boldsymbol{\mathcal{I}}_1$) *For every* $j = 1, \ldots, K$ *and* $A \subsetneq A_j \setminus A_{j-1}$, $\exists J \in \mathcal{P}_{A_j} \setminus \mathcal{P}_{A\cup A_{j-1}}, J' \in \mathcal{P}_{A\cup A_{j-1}} \setminus \mathcal{P}_{A_{j-1}}$ *such that* $(J \cap J') \setminus A_{j-1} \neq \emptyset$,

$\boldsymbol{\mathcal{I}}_2$) *For all* $j = 2, \ldots, K$ *there exists* $s \in A_{j-1} \setminus A_{j-2}$, $J \in \mathcal{P}_{A_j} \setminus \mathcal{P}_{A_{j-1}}$ *such that* $J \ni s$.

(We will come back to these perhaps opaque conditions below.)

For a Poisson Point Process, PPP for short, $\sum \delta_{\eta_i}$ of density $xt^{-x-1}dt$ on $\mathbb{R}^+$ for some $x \in (0,1)$, we understand by $\sum \delta_{\overline{\eta}_i}$ the normalized process where $\overline{\eta}_i = \eta_i / \sum_j \eta_j$, and denote by P_x its law.

Our main result is to determine the weak limits of the measures $\Xi_{\beta,N}$ and $\Xi^{(m)}_{\beta,N}$ describing at the same time the limiting (marginal) Gibbs distribution, and the limiting overlap structure.

Theorem 1.4. *Let the irreducible hamiltonian be given, with chain* $\mathbf{T} = (A_0, A_1, \ldots, A_K)$ *and associated sequence of phase transitions* $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_K)$. *Define the times through* $t_j = \log(x_K/x_{K-j})$, $x_j = x_j(\beta) = \beta_j/\beta$ *and set* $\mathbf{t} = \{t_0, \ldots, t_K\}$.

a) *(Low temperature) For* $\beta > \beta_K$,

$$\lim_{N\to\infty} \Xi_{N,\beta} = P_{x_K} \sqcap Q_{\mathbf{T},\mathbf{t}}$$

weakly.

b) *(Pure states) Let* $m < K$ *and set* $\mathbf{t}^{(m)} = \{t_0, \ldots, t_m\}$. *For* $\beta > \beta_m$

$$\lim_{N\to\infty} \Xi^{(m)}_{N,\beta} = P_{x_m} \sqcap Q_{\mathbf{T}^{(m)},\mathbf{t}^{(m)}}$$

weakly.

According to Theorem 1.4, the only possible "marks" in the $N \to \infty$ limit are the ones from the chain $\mathbf{T}$. This is the onset ultrametricity.

Remark also that the hierarchical structure does not enter into the law P_{x_K} of the limiting Gibbs weights, which is even independent of the overlaps. As for the β-dependence of these objects, there is only a slight difference showing up for β strictly less than the last phase transition β_K (claim b): in this regime, individual configurations are negligible in the thermodynamical limit, and in order to catch a "macroscopic" weight, we have to lump together exponentially many of the σ's. We can achieve this by simply taking the collections of pure states $E_\tau \stackrel{\text{def}}{=} \{\sigma \in \Sigma_N : \ \sigma_{A_m} = \tau\}$ where $\tau \in \Sigma_{N,A_m}$. The Gibbs weights $\mathcal{G}^{(m)}_{\beta,N}(E_\tau)$ of these disjoint sets remain macroscopic in the limit. Lastly, remark that there is no transition for $\beta \uparrow \infty$ in the behavior of the marginal.

We next explain heuristically the irreducibility conditions.

Condition $\mathcal{I}_1$ and suppression of structures.

We consider here three models with limiting free energy of a REM, which nevertheless display strikingly different microscopic behavior.

a) First we take a GREM with two branches (the prototypical example of a priori hierarchical model) and thus $\mathcal{P} = \big\{\{1\}, \{1,2\}\big\}$ with parameters chosen so that the chain is $\mathbf{T} = \{1, 2\}$. It is not difficult to see that there exist constants $a_{N,1}$ and $a_{N,2}$ growing linearly with the size of the system such that (putting $a_N = a_{N,1} + a_{N,2}$) the collection of points $(X_\sigma - a_N;\ \sigma \in \Sigma_N)$ converges to a Poisson point process. In some sense, the random variables $X^{\{1\}}_{\sigma_1}$ and $X^{\{1,2\}}_{\sigma_1,\sigma_2}$ are somehow *lumped together.* On a rigorous level, things are quite trivial, since one can fully exploit the a priori hierarchical structure, be it through a *Laplace transform* [**4**] or a *Moment*

Generating [**6**] approach. None of these carry over when the system lacks a priori hierarchical organization, the somewhat default situation for spin glasses of the form (1.1). There is however a simple feature which does; consider

$$\Omega_{\{1\}} \stackrel{\text{def}}{=} \Big\{ \exists\, \sigma,\tau \in \Sigma_N :\; X^{\{1\}}_{\sigma_1}, X^{\{1\}}_{\tau_1} \approx a_{N,1},\; X^{\{1,2\}}_{\sigma_1,\sigma_2}, X^{\{1,2\}}_{\tau_1,\tau_2} \approx a_{N,2},\; q(\sigma,\tau) = \{1\} \Big\} \tag{1.16}$$

The point here is that the $\mathbb{P}$-probability of the event $\Omega_{\{1\}}$ is in the limit $N \to \infty$ vanishing. This also holds if we require $q(\sigma,\tau) = \{2\}$. Inside the (unique) level of the tree no particular structure is present: either relevant configurations differ on both spins (in which case the associated random variables X_σ, X_τ are independent) or they coincide. This explains the REM-like behavior.

b) Consider the hamiltonian on $\mathcal{P} = \big\{\{1\},\{2\},\{1,2\}\big\}$ with $\mathbf{T} = \{1,2\}$. For finite N, no particular organization is present. Yet, exactly as in case **a)**, for N large enough and on a set of $\mathbb{P}$-probability close to one, given two relevant configurations $\sigma,\tau \in \Sigma_N$, $\sigma_1 = \tau_1$ implies $\sigma_2 = \tau_2$ and the other way around: this kind of non hierarchical dependencies is also suppressed. The overlap is either maximal or the empty set. This is **not** always the case.

c) For $\mathcal{P} = \big\{\{1\},\{2\}\big\}$ with $\mathbf{T} = \{1,2\}$, the above considerations break down: with non vanishing probability, one can find relevant $\sigma,\tau,\tau' \in \Sigma_N$ such that $q(\sigma,\tau) = \{1\}$ and $q(\sigma,\tau') = \{2\}$, showing that this kind of non hierarchical dependencies cannot be suppressed. This is in some sense due to the "missing bond" linking the two indeces. Although it does not prevent the system to display some kind of clustering on the level of the free energy, it does have a dramatic impact on the behavior of the Gibbs measure, and in fact, such a system cannot be described in the limit by Ruelle's probability cascades.

Condition $\mathcal{I}_2$ and propagation along the tree.

d) Consider a two-levels GREM as in **a)** but underlying parameters chosen in such a way that $\mathbf{T} = \big\{\{1\},\{1,2\}\big\}$. This time it is easy to see that the probability that there exist relevant configurations $\sigma,\tau \in \Sigma_N$ such that $q(\sigma,\tau) = 2$ is vanishing, but not if we require $q(\sigma,\tau) = 1$: given that $\sigma,\tau \in \Sigma_N$ coincide on the second index ($\sigma_2 = \tau_2$) then automatically on the first as well, in which case the two configurations coincide. The situation is fully analogous in the non hierarchical model of case b) but parameters tuned in so that that $\mathbf{T} = \big\{\{1\},\{1,2\}\big\}$.

e) Finally, let $\mathcal{P} = \big\{\{1\},\{2\},\{2,3\}\big\}$ and $\mathbf{T} = \big\{\{1\},\{1,2,3\}\big\}$. In this case, also on the finer level there is *clustering* on the second level (e.g. $\sigma_2 = \tau_2$ implies $\sigma_3 = \tau_3$), but it is not true that $\sigma_2 = \tau_2$ implies $\sigma_1 = \tau_1$ nor $\sigma_3 = \tau_3$ implies $\sigma_1 = \tau_1$. Intuitively, the lack of a "linking bond" from the second branch to the first prevents the coincidence of the spins indexed by A_2 to propagate "upwards" to the spins indexed by A_1.

The above considerations on the irreducibility conditions already contain some of the crucial ideas for the proof to Theorem 1.4, which we now sketch for the readers convenience. In words, the strategy is essentially that of a *pointwise* analog of the Second Moment Method, enabling to single out with relatively little effort ultrametric configurations as the only possible winners of the energy/entropy competition. Restricted to the particular sub-class of hierarchical models (GREM), the point of view taken here provides quite an alternative solution to that presented in [**6**], and additionally clarifies the mechanism pertaining to the *coarsening* of the initial, a priori ultrametric.

We emphasize that we think of the irreducible spin glass $(X_\sigma; \sigma \in \Sigma_N)$ with associated chain $\mathbf{T} = (A_0, A_1, \dots, A_K)$ as being given. Unless otherwise stated, any claim will refer to this hamiltonian. Regrettably, we still need an arsenal of notational conventions.

Generalities. For $\tau \in \Sigma_{N,J}$, and $J' \subset J$ we write $\tau_{J'}$ for the projected configuration $(\tau_j;\ j \in J')$. For $j = 1 \dots K$ we write

$$\Delta_j \stackrel{\text{def}}{=} \alpha(A_j) - \alpha(A_{j-1}), \quad G_j \stackrel{\text{def}}{=} \gamma(A_j) - \gamma(A_{j-1}),$$

and for a subset $A \subsetneq A_j \setminus A_{j-1}$,

$$\widehat{\mathcal{P}}_{A,j} \stackrel{\text{def}}{=} \mathcal{P}_{A \cup A_{j-1}} \setminus \mathcal{P}_{A_{j-1}},\ \widehat{\mathcal{P}}^c_{A,j} \stackrel{\text{def}}{=} \mathcal{P}_{A_j} \setminus \widehat{\mathcal{P}}_{A \cup A_{j-1}},\ \widehat{\alpha}_j(A) \stackrel{\text{def}}{=} \alpha(A \cup A_{j-1}) - \alpha(A_{j-1})$$

as well as $\widehat{\alpha}^c_j(A) \stackrel{\text{def}}{=} \Delta_j - \widehat{\alpha}_j(A)$. An important rôle in the whole analysis is played by the constants

$$a_{N,j}(A) \stackrel{\text{def}}{=} \beta_j \widehat{\alpha}_j(A) N - \frac{1}{2\beta_j} \log N + \frac{1}{\beta_j} \log \beta_j \sqrt{2\pi \widehat{\alpha}_j(A)},\ a_{N,j} \stackrel{\text{def}}{=} a_{N,j}(A_j \setminus A_{j-1}).$$

For a multi-index $\mathbf{i} \stackrel{\text{def}}{=} (i_1, \dots, i_j) \in \mathbb{N}^j$ and $k < j$, we write $\mathbf{i}_k = (i_1, \dots, i_k)$ for its restriction to the first k indeces.

Random variables. By $(Y_J, J \in \mathcal{P})$ we denote a family of independent centered gaussians, $\mathbb{E}\big(Y_J^2\big) = a_J$, and shorten notations by setting

$$Y_j \stackrel{\text{def}}{=} \sum_{J \in \mathcal{P}_{A_j} \setminus \mathcal{P}_{A_{j-1}}} Y_J, \quad \overline{Y}_j \stackrel{\text{def}}{=} \sqrt{N} Y_j - a_{N,j}, \quad \widehat{Y}_j \stackrel{\text{def}}{=} \sum_{l=1,\dots,j} \overline{Y}_l,$$

$$Y_{j,A} \stackrel{\text{def}}{=} \sum_{J \in \widehat{\mathcal{P}}_{A,j}} Y_J, \quad Y^c_{j,A} \stackrel{\text{def}}{=} \sum_{J \in \widehat{\mathcal{P}}^c_{A,j}} Y_J.$$

By (Z_J) we denote an independent copy, with same distribution as the $Y's$, and write in full analogy $Z_{j,A}, Z^c_{j,A}, \overline{Z}_j, \widehat{Z}_j$.

For $\sigma \in \Sigma_{N,A_j}$ we write $\sigma = (\sigma(1), \ldots, \sigma(j))$ with $\sigma(k) = (\sigma_i; i \in A_k \setminus A_{k-1})$ and

$$X_\sigma = \sum_{j=1}^{K} X_{\sigma(1),\ldots,\sigma(j)}, \quad X_{\sigma(1),\ldots,\sigma(j)} \stackrel{\text{def}}{=} \sum_{J \in \mathcal{P}_{A_j} \setminus \mathcal{P}_{A_{j-1}}} X^J_{\sigma_J}$$

$$\overline{X}_{\sigma(1),\ldots,\sigma(j)} \stackrel{\text{def}}{=} X_{\sigma(1),\ldots,\sigma(j)} - a_{N,j}, \quad \widehat{X}_{\sigma(1),\ldots,\sigma(j)} \stackrel{\text{def}}{=} \sum_{l=1}^{j} \overline{X}_{\sigma(1),\ldots,\sigma(l)}.$$

Critical subsets. According to the construction of the chain $\mathbf{T}$, there may exist strict subsets $A \subsetneq A_j \setminus A_{j-1}$ for some $j = 1, \ldots, K$ such that

$$\rho(A_{j-1}, A \cup A_{j-1}) = \beta_j \left(\text{that is } \gamma(A_j) - \gamma(A_{j-1} \cup A) \Big/ \widehat{\alpha}_j(A) = \beta_j^2 / 2\log 2\right),$$

in which case we call the subsets *critical*.

Ultrametricity. We say that $\sigma, \tau \in \Sigma_{N,A_j}$ (for some $j = 1, \ldots, k$) form a *non ultrametric couple* if there exists $k = 1, \ldots, j$ and $s \in A_k \setminus A_{k-1}$ such that $\sigma_s = \tau_s$ but $\sigma_{A_k} \neq \tau_{A_k}$ (i.e. for some $r \in A_k$ it still holds $\sigma_r \neq \tau_r$).

Constants. We denote by *const* a strictly positive constant, not necessarily the same at different occurences. For $X, Y > 0$ we write $X \lesssim Y$ if $X \leq \mathit{const} \times Y$ (for sequences: $X_N \lesssim Y_N$ stands for $X_N \leq \mathit{const} \times Y_N$ for $N \geq N_o$ for some $N_o \in \mathbb{N}$).

For the proof of Theorem 1.4 it will be of crucial importance to obtain a good control of the *energy levels*: For given $j = 1, \ldots, K$, consider the collection

$$\left(\widehat{X}_{\sigma(1),\ldots,\sigma(j)}; \sigma \in \Sigma_{N,A_j}\right),$$

the process of the energy levels corresponding then to the choice $j = K$. A fixed realization induces an element $\mathcal{N}_{N,j} \in \mathcal{M}_{mp}\left(\mathbb{R}^{(2)} \times 2^{A_j}\right)$ by taking

$$\sum_{\sigma,\tau \in \Sigma_{N,A_j}} \delta_{\{\widehat{X}_{\sigma(1),\ldots,\sigma(j)}, \widehat{X}_{\tau(1),\ldots,\tau(j)}; q(\sigma,\tau)\}}.$$

We denote by $\widehat{X}_{N,j}$ its law.

We will have to consider a number of objects, usually obtained from the process of the energy levels through a truncation procedure, in which case it is notationally more convenient to think of the new process as living not on the original space of configurations but rather on a *random subset*. As a first, illustrative example, let us consider Σ_{N,A_j} and retain its configurations σ only such that $\overline{X}_{\sigma(1),\ldots,\sigma(k)} \in [-R,R]$ for some $R>0$ and every $k \leq j$. This generates a random space of configurations which we then call Σ^R_{N,A_j}. A fixed realization $\left(\widehat{X}_{\sigma(1),\ldots,\sigma(j)};\ \sigma \in \Sigma^R_{N,A_j}\right)$, that is

$$\left(\widehat{X}_{\sigma(1),\ldots,\sigma(j)};\ \sigma \in \Sigma_{N,A_j}, \forall\, k \leq j\ \overline{X}_{\sigma(1),\ldots,\sigma(k)} \in [-R,R]\right),$$

induces naturally an element $\mathcal{N}^R_{N,j} \in \mathcal{M}_{mp}(\mathbb{R}^{(2)} \times 2^{A_j})$ whose law is denoted $\widehat{X}^R_{N,j}$.

As a preliminary motivation to the thinning comes the following:

Proposition 1.5. *Let $\mathfrak{M} \subset \mathbb{R}$ be a bounded real subset. To $\epsilon > 0$ there exists finite R and $N_o = N_o(\epsilon)$ such that*

$$\mathbb{P}\Big[\exists \tau \in \Sigma_{N,A_j} \setminus \Sigma^R_{N,A_j} : \widehat{X}_{\tau(1),\ldots,\tau(j)} \in \mathfrak{M}\Big] \leq \epsilon.$$

for $N \geq N_o$.

Since weak convergence of $\widehat{X}_{N,j}$ is equivalent to convergence of the finite dimensional distributions, the Proposition 1.5 is quite useful. To see this, consider a finite subset $\mathcal{J} \subset \mathbb{N}$, a set of overlaps $Q \subset 2^I$, as well as a finite collection of bounded real intervals $\mathcal{M}$. To $\epsilon > 0$ we may find R large enough such that

$$\begin{aligned}
&\Bigg|\mathbb{P}\Big[\mathcal{N}_{N,j}(\mathfrak{m},\mathfrak{m}';q) = k,\ \text{for } \mathfrak{m},\mathfrak{m}' \in \mathcal{M}, q \in Q, k \in \mathcal{J}\Big] \\
&\qquad\qquad - \mathbb{P}\Big[\mathcal{N}^R_{N,j}(\mathfrak{m},\mathfrak{m}';q) = k,\ \text{for } \mathfrak{m},\mathfrak{m}' \in \mathcal{M}, q \in Q, k \in \mathcal{J}\Big]\Bigg| \\
&\leq \mathbb{P}\left[\exists \tau \in \Sigma_{N,A_j} \setminus \Sigma^R_{N,A_j} : \widehat{X}_{\tau(1),\ldots,\tau(j)} \in \bigcup_{\mathfrak{m}\in\mathcal{M}} \mathfrak{m}\right] \leq \epsilon.
\end{aligned} \tag{1.17}$$

Therefore, in order to study the weak limit of $\mathcal{N}_{N,j}$, it suffices to study the weak limit of $\mathcal{N}^R_{N,j}$, the process defined on the thinned space Σ^R_{N,A_j}.

We proceed to further thinnings.

The first is by means of a truncation procedure which is innerly related to condition $\boldsymbol{\mathcal{I}}_1$, cfr. Remark 1.10 below. For $\varepsilon_1 > 0, k = 1,\ldots,j$ and critical subset $A \subsetneq A_k \setminus A_{k-1}$ we say that $\mathbf{T}_1(\sigma,k,A,\varepsilon_1)$ holds if

$$\frac{1}{\widehat{\alpha}_k(A)} \sum_{J\in\widehat{\mathcal{P}}_{A,k}} X^J_{\sigma_J} - \frac{1}{\widehat{\alpha}^c_k(A)} \sum_{J\in\widehat{\mathcal{P}}^c_{A,k}} X^J_{\sigma_J} \leq -\varepsilon_1\sqrt{N}.$$

To simplify notations we may also say that $\mathbf{T_1}(\varepsilon_1)$ holds, tacitly understanding that it holds for all critical subsets. Moreover, for $\varepsilon_2 > 0, k = 1, \ldots, j$ and (critical and non critical) subsets $A \subsetneq A_k \setminus A_{k-1}$ such that $\widehat{\alpha}_k(A) > 0$, we say that $\mathbf{T}_2(\sigma, k, A, \varepsilon_2)$ holds if

$$\sum_{J \in \widehat{\mathcal{P}}_{A,k}} X^J_{\sigma_J} \leq \beta_k \widehat{\alpha}_k(A)(1+\varepsilon_2)N.$$

Again, $\mathbf{T}_2(\varepsilon_2)$ holds, if it holds for all possible subsets. The random space obtained by retaining those $\sigma \in \Sigma^R_{N,A_j}$ only such that $\mathbf{T_1}(\varepsilon_\mathbf{1})$ and $\mathbf{T_2}(\varepsilon_2)$ hold is denoted by $\Sigma^{R,\varepsilon_1,\varepsilon_2}_{N,A_j}$. We also introduce the process $\mathcal{N}^{R,\varepsilon_1,\varepsilon_2}_{N,j} \in \mathcal{M}_{mp}\big(\mathbb{R}^{(2)} \times 2^{A_j}\big)$ naturally induced by the collection $(\widehat{X}_{\sigma(1),\ldots,\sigma(j)}; \sigma \in \Sigma^{R,\varepsilon_1,\varepsilon_2}_{N,A_j})$ and denote by $\widehat{X}^{R,\varepsilon_1,\varepsilon_2}_{N,j}$ its law.

Proposition 1.6. *Let $R, \varepsilon_2 > 0$. Then,* $\lim_{\varepsilon_1 \downarrow 0} \lim_{N \uparrow \infty} \mathbb{P}\Big[\Sigma^R_{N,A_j} \setminus \Sigma^{R,\varepsilon_1,\varepsilon_2}_{N,A_j} \neq \emptyset\Big] = 0.$

This result has an important consequence. We first write

$$\mathcal{N}^R_{N,j} = \mathcal{N}^{R,\varepsilon_1,\varepsilon_2}_{N,j} + (\mathcal{N}^R_{N,j} - \mathcal{N}^{R,\varepsilon_1,\varepsilon_2}_{N,j}),$$

and assume for the time being that $\widehat{X}^{R,\varepsilon_1,\varepsilon_2}_{N,j}$ converges for $N \to \infty$ to a weak limit. Since

$$\big\{\big(\mathcal{N}^R_{N,j} - \mathcal{N}^{R,\varepsilon}_{N,j}\big)(\mathbb{R}, \mathbb{R}; \emptyset) > 0\big\} \subset \big\{\Sigma^R_{N,A_j} \setminus \Sigma^{R,\varepsilon_1,\varepsilon_2}_{N,A_j} \neq \emptyset\big\},$$

Proposition 1.6 implies that the weak limit should not depend on ε_2, with the contribution of the process $(\mathcal{N}^R_{N,j} - \mathcal{N}^{R,\varepsilon_1,\varepsilon_2}_{N,j})$ being asymptotically irrelevant. Assuming furthermore that the weak limit behaves well in ε_1 (we will make this precise) it must be that $\lim_{N\uparrow\infty} \widehat{X}^R_{N,j} = \lim_{\varepsilon_1\downarrow 0} \lim_{N\uparrow\infty} \widehat{X}^{R,\varepsilon_1,\varepsilon_2}_{N,j}$.

The above considerations are indeed correct: for $\varepsilon_2 > 0$ "small enough" the law $\widehat{X}^{R,\varepsilon_1,\varepsilon_2}_{N,j}$ does converge weakly to a well behaved limit. (Here and in what follows "small enough" means that there exists strictly positive $\overline{\varepsilon}_2$ depending on the underlying parameters only, cfr. (1.49) and (1.53) below, such that for $\varepsilon_2 < \overline{\varepsilon}_2$ weak convergence holds.)

We next describe the weak limit. Consider a PP $(x_\mathbf{i}; \mathbf{i} \in \mathbb{N}^j)$ where $x_\mathbf{i} = x^1_{\mathbf{i}_1} + \cdots + x^j_{\mathbf{i}_j}$ with the following properties:

- For $l = 1, \ldots, j$ and multi-index $\mathbf{i}_{l-1}$, the point process $(x^l_{\mathbf{i}_{l-1}, i_l};\ i_l \in \mathbb{N})$ is poissonian with density $\mathcal{C}_{l,\varepsilon_1} \cdot \beta_l e^{-\beta_l t} dt$ on $[-R, R]$ and 0 otherwise.
- The x^l are independent for different l.
- $(x^l_{\mathbf{i}_{l-1}, i_l};\ i_l \in \mathbb{N})$ are independent for different $\mathbf{i}_{l-1}$.
- If $A_l \setminus A_{l-1}$ contains no critical subsets, then $\mathcal{C}_{l,\varepsilon_1} = 1$. Otherwise

$$\mathcal{C}_{l,\varepsilon_1} = \mathbb{P}\left[\left\{\frac{Y_{l,A}}{\widehat{\alpha}_l(A)} - \frac{Y^c_{l,A}}{\widehat{\alpha}^c_l(A)} \leq -\varepsilon_1\right\} \forall A \subsetneq A_l \setminus A_{l-1},\ A \text{ is critical}\right].$$

Given two points $x_\mathbf{i}$ and $x_{\mathbf{i}'}$, we define their overlap $q_{\mathbf{i},\mathbf{i}'}$ to be A_m where $m = \max\big\{l \leq j\ :\ \mathbf{i}_l = \mathbf{i}'_l\big\}$. A fixed realization of the PP induces naturally an element $\mathcal{N}^{R,\varepsilon_1}_j \in \mathcal{M}_{mp}\big(\mathbb{R}^{(2)} \times 2^{A_j}\big)$ whose law is denoted X^{R,ε_1}_j.

Proposition 1.7. *For $\varepsilon_1 > 0$ and small enough $\varepsilon_2 > 0$, $\lim_{N\to\infty} \widehat{X}_{j,N}^{R,\varepsilon_1,\varepsilon_2} = \widehat{X}_j^{R,\varepsilon_1}$ weakly.*

A crucial ingredient in the proof of Proposition 1.7 is the following

Proposition 1.8. *For $\varepsilon_1 > 0$ and small enough $\varepsilon_2 > 0$,*

$$\lim_{N\to\infty} \mathbb{P}\Big[\exists \text{ non ultrametric couples in } \Sigma_{N,A_j}^{R,\varepsilon_1,\varepsilon_2}\Big] = 0.$$

We sketch the strategy of the proof of these two results. As for Proposition 1.8, we show in a first step that on a set of probability subexponentially close (in the size of the system) to unity, for two configurations $\sigma, \tau \in \Sigma_{N,A_j}^{R,\varepsilon_1,\varepsilon_2}$, coincidence of spins on some index $i \in A_k \setminus A_{k-1}$ (for some $k = 1, \dots, j$) implies that $\sigma_r = \tau_r$ for all $r \in A_k \setminus A_{k-1}$. In a second step we show that coincidence of spins on the level $A_k \setminus A_{k-1}$ propagates backward to the root: on a set of probability arbitrarily close to unity, given $\sigma, \tau \in \Sigma_{N,A_j}^{R,\varepsilon_1,\varepsilon_2}$ such that $\sigma(k) = \tau(k)$, then automatically $\sigma(i) = \tau(i)$ for all $i = 1, \dots, k$. Putting these two steps together we see that the overlap of two configurations on $\Sigma_{N,A_j}^{R,\varepsilon_1,\varepsilon_2}$ must belong to the chain $\mathbf{T}^j = \{\emptyset, A_1, \dots, A_j\}$.

The main idea in the proof of Proposition 1.7 is to work inductively on the level j, thereby exploiting the properties of the point processes involved in the limiting picture. These are such that once we know what happens on level $j-1$ (the distribution on the real axis of the points $x_{i_1}^1 + \cdots + x_{i_1,\dots,i_{j-1}}^{j-1}$, as well as their overlap structure) the "full process" is obtained by adding random points independently: no matter what the (hierarchical) overlap structure, given $k \in \mathbb{N}$ multi-indeces $\mathbf{i}^1, \dots, \mathbf{i}^k \in \mathbb{N}^{j-1}$, and $B_1, \dots, B_k \subset (-R, R)$, we have the following equality in distribution

$$\left(\sum_{l\in\mathbb{N}} \delta_{x^j_{\mathbf{i}^1,l}}(B_1), \dots, \sum_{l\in\mathbb{N}} \delta_{x^j_{\mathbf{i}^k,l}}(B_k)\right) \overset{(d)}{=} \Big(V_1, \dots, V_k\Big) \tag{1.18}$$

with the random variables $V_r, r = 1, \dots, k$ being independent, Poisson-distributed of parameters $\mu_{\varepsilon_1}(B_r) \overset{\text{def}}{=} \int_{B_r} C_{j,\varepsilon_1}\beta_j e^{-\beta_j t} dt$. Clearly, the problem of computing a given finite dimensional distribution of the limiting process $\widehat{X}_j^{R,\varepsilon_1}$ can be brought back (by conditioning) to the problem of computing the distribution of a random vector such as (1.18). As for the finite N system, Proposition 1.8 allows to rule out, on every level, events not satisfying ultrametricity. So, the main task will be to prove that, for given family of reference configurations $\sigma^1, \dots, \sigma^k \in \Sigma_{N,A_{j-1}}$ with a certain overlap structure $q(\sigma^r, \sigma^t) \in \{\emptyset, \dots, A_{j-1}\}$, and $r, s = 1, \dots, k$ the distribution of the random vector

$$\left(\sum^{(1)} \delta_{\overline{X}_{\sigma(1),\dots,\sigma(j)}}(B_1), \dots, \sum^{(k)} \delta_{\overline{X}_{\sigma(1),\dots,\sigma(j)}}(B_k)\right) \tag{1.19}$$

(with the sum in the rth-term running over those $\sigma \in \Sigma_{N,A_j}$ such that $\sigma_{A_{j-1}} = \sigma^r$ and conditions $\mathbf{T_1}(\varepsilon_1)$ satisfied for all critical subsets in $A_j \setminus A_{j-1}$ and $\mathbf{T_2}(\varepsilon_2)$ satisfied for all non-critical subsets) is approximately multivariate Poisson, cfr. (1.18). To prove this, we will resort to the *Chen-Stein method*, a particularly efficient tool in Poisson approximation.

Theorem 1.9. *Let $\widehat{X}_j$ be the law of the process on $\mathcal{M}_{mp}(\mathbb{R}^{(2)} \times 2^{A_j})$ naturally induced by a PP $(y_{\mathbf{i}}, \mathbf{i} \in \mathbb{N}^j)$ with the following properties: for $l = 1, \dots, j$ and multi-index $\mathbf{i}_{l-1}$, the point process $(y^l_{\mathbf{i}_{l-1}, i_l};\ i_l \in \mathbb{N})$ is poissonian with density $\mathcal{C}_l \cdot \beta_l e^{-\beta_l t} dt$ on $\mathbb{R}$; the y^l are independent for different l; $(y^l_{\mathbf{i}_{l-1}, i_l};\ i_l \in \mathbb{N})$ are independent for different $\mathbf{i}_{l-1}$; if $A_l \setminus A_{l-1}$ contains no critical subsets, then $\mathcal{C}_l = 1$. Otherwise*

$$\mathcal{C}_l = \mathbb{P}\left[\left\{ \frac{Y_{l,A}}{\widehat{\alpha}_l(A)} - \frac{Y^c_{l,A}}{\widehat{\alpha}^c_l(A)} \leq 0 \right\} \forall A \subsetneq A_l \setminus A_{l-1},\ A \text{ is critical} \right].$$

The following then holds:

i) *$\widehat{X}_{j,N}$ converges weakly to $\widehat{X}_j$.*
ii) *The point process of extremes $\sum_{\sigma \in \Sigma_{N,A_j}} \delta_{\widehat{X}_{\sigma(1),\dots,\sigma(j)}}$ converges weakly to the point process $\sum_{i_1,\dots,i_j \in \mathbb{N}} \delta_{y^1_{i_1} + \dots y^j_{i_1,\dots,i_j}}$.*

Proof. According to (1.17), the finite dimensional distributions of $\widehat{X}_{j,N}$ can be approximated up to arbitrarily small error by the finite dimensional distributions of $\widehat{X}^R_{j,N}$. By the considerations following Proposition 1.6, the finite dimensional distribution of the latter can be approximated by the finite dimensional distributions of $\widehat{X}^{R,\varepsilon_1,\varepsilon_2}_{j,N}$ (up to an error term which vanishes in the limit $N \to \infty, \varepsilon_1 \to 0$). But $\widehat{X}^{R,\varepsilon_1,\varepsilon_2}_{j,N}$ converges weakly, at least for small enough ε_2, to $\widehat{X}^{R,\varepsilon_1}_j$ and so do the finite dimensional distributions. Quite plainly, $\lim_{R\uparrow\infty} \lim_{\varepsilon_1 \downarrow 0} \widehat{X}^{R,\varepsilon_1}_j = \widehat{X}_j$. This settles claim *i*), which then automatically implies *ii*). □

Remark 1.10. *The first irreducibility condition enters crucially in the truncation procedure, a feature which is inherited by the constants $\mathcal{C}$. In fact, $\mathbf{T}_1$ makes sense only provided $\boldsymbol{\mathcal{I}}_1$ is satisfied, automatically ensuring that $\mathcal{C}_l > 0$. We explain this directly on the limiting constant: for critical $A \subsetneq A_l \setminus A_{l-1}$, by simple properties of real numbers we also have*

$$\Big[\gamma(A_l) - \gamma(A \cup A_{l-1})\Big] \Big/ \widehat{\alpha}^c_l(A) = \beta_j^2 / (2 \log 2).$$

By $\boldsymbol{\mathcal{I}}_1$ there exists $J \in \mathcal{P}_{A_l} \setminus \mathcal{P}_{A \cup A_{l-1}}$ with $J \cap A \neq \emptyset$, in which case $\widehat{\alpha}^c_l(A) > \widehat{\alpha}_l\big(A_l \setminus (A \cup A_{l-1})\big)$. This implies that the relative complement $A_l \setminus (A \cup A_{l-1})$ cannot be critical,

$$\Big[\gamma(A_l) - \gamma(A \cup A_{l-1})\Big] \Big/ \widehat{\alpha}_l\Big(A_l \setminus (A \cup A_{l-1})\Big) > \beta_j^2 / 2 \log 2.$$

*To further clarify, consider the example $X_\sigma = X^{\{1\}}_{\sigma_1} + X^{\{2\}}_{\sigma_2}$ with parameters $a_1 = a_2 = \gamma_1 = \gamma_2 = 1/2$. The associated chain is then $\mathbf{T} = \{A_o = \emptyset, A_1 = \{1,2\}\}$ and both subsets $\{1\}, \{2\}$ are critical. Evidently, **C1** does not hold. Also, condition $\mathbf{T}_1$ is (to given ε) meaningless since it is fulfilled by those $\sigma \in \Sigma_N$ such that $X^{\{1\}}_{\sigma_1} - X^{\{2\}}_{\sigma_2} \leq -\varepsilon\sqrt{N}$ and simultaneously $X^{\{2\}}_{\sigma_2} - X^{\{1\}}_{\sigma_1} \leq -\varepsilon\sqrt{N}$: there is no such configuration.*

Remark 1.11. *There is an interesting interpretation of the critical constants for the case of the GREM. To see this, consider on an additional probability space* $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ *a Brownian Bridge* $(\mathcal{B}(t), 0 \le t \le 1)$*, starting and ending in* 0*. The a priori hierarchical structure of the GREM determines in some sense a particular "time arrow", reflected in the nestling of the critical subset,* $A_1^{crit} \subsetneq A_2^{crit}, \ldots, A_j^{crit} \subsetneq A_l \setminus A_{l-1}$*. Defining the "times"* $s_r = \widehat{\alpha}_l(A_r^{crit})$*, for* $r = 1, \ldots, j$ *one can show that the critical constants are given by* $\mathcal{C}_l = \tilde{\mathbb{P}}\Big[\mathcal{B}(s_1) \le 0, \ldots, \mathcal{B}(s_j) \le 0\Big]$*. This is by no means fortuitous; there is in fact a tight link between the issues addressed in this work and those related to precise second-order corrections of the maximal displacement of branching brownian motion* [**7**]*. Contrary to the GREM, there is no "Brownian bridge representation" of the critical constants for genuinely non hierarchical hamiltonians.*

So far we have addressed the issue from the point of view of the energy levels only, with the implicit motivation that Gibbs measure is after all a "functional" of the latter; so what one still needs are some kind of "compactness arguments", which we next describe. The next Proposition is valid in the range $\beta \in (\beta_m, \beta_{m+1})$ for arbitrary $m = 1, \ldots, K$.

Proposition 1.12. *Let* $\epsilon > 0$*. There exists* $C > 0$ *such that for large enough* N *it holds*

$$\mathbb{P}\left[\mathcal{G}_{\beta,N}\left(\exists j \le m : \overline{X}_{\sigma(1),\ldots,\sigma(j)} \notin (-C, C)\right) \ge \epsilon\right] \le \epsilon.$$

An important ingredient in the proof of Proposition 1.12 is Lemma 1.13 below, whose motivation goes as follows: according to the limiting free energy in the range $\beta_m < \beta < \beta_{m+1}$ for m strictly less than K, a partial structure only has emerged. A portion of the system is *frozen* and displays hierarchical organization (the collection of points given by $\widehat{X}_{\sigma(1),\ldots,\sigma(m)}, \sigma \in \Sigma_{N,A_m}$), with the crucial property that relevant configurations coincide up to a certain level $i = 0, \ldots, m$ of the chain, and then differ overall on the levels $i+1, \ldots, m$. For $m+1, \ldots, K$ (corresponding to the portion of the system in *high-temperature*) there is no such geometrical structure; however, this does not have a dramatic impact on the global properties of the system, mainly because the fluctuations of the involved random variables are in the limit $N \to \infty$ negligible. To be more precise, fix $\sigma \in \Sigma_{N,A_m}$ and set

$$Z_\sigma \stackrel{\text{def}}{=} \sum_{\tau \in \Sigma_N : \tau_{A_m} = \sigma} \exp\left[\beta\Big(X_{\tau(1),\ldots,\tau(m+1)} + \cdots + X_{\tau(1),\ldots,\tau(K)}\Big)\right].$$

Lemma 1.13. *Let* $\beta_m < \beta < \beta_{m+1}$*. There exist constants* $\delta_1, \delta_2 \in (0,1)$ *such that*

$$\mathbb{P}\left[\Big|\log \frac{Z_\sigma}{\mathbb{E}[Z_\sigma]}\Big| \ge N^{-\delta_1}\right] \lesssim \exp\left[-N^{\delta_2}\right].$$

The rest of the chapter is devoted to the proofs of the main results. In Section 1.2 we prove Theorem 1.1: in a first step, Section 1.2.1, we obtain the limiting free energy of the non hierarchical GREMs in terms of a variational principle, which we then solve in Section 1.2.2. We then address in Section 1.3 the Gibbs measure. Proposition 1.5 is proved in Section 1.3.1, Proposition 1.6 is proved in section 1.3.2 while Proposition 1.8 is proved in Section 1.3.3. This allows then to tackle the proof of Proposition 1.7, which is given in Section 1.3.4. In Section 1.3.5 we give the proof of Lemma 1.13 and Proposition 1.3. Finally, we Theorem 1.4 is proved in Section 1.3.5.

1.2. Free energy of the non hierarchical GREM, Proofs

1.2.1. Second Moment estimates. We fix some notations: If $(a_N)_{N\in\mathbb{N}}$ and $(b_N)_{N\in\mathbb{N}}$ are two sequences of positive real numbers, we write $a_N \asymp b_N$ if for all $\varepsilon > 0$ there exists $N_0(\varepsilon) \in \mathbb{N}$ such that

$$\mathrm{e}^{-\varepsilon N} b_N \leq a_N \leq \mathrm{e}^{-\varepsilon N} b_N$$

for $N \geq N_0$. We also write $a_N \ll b_N$ if for some $\delta > 0$, one has $a_N \leq b_N \mathrm{e}^{-\delta N}$, again for large enough N. In that case, we also write $a_N = \Omega(b_N)$. The same notations are used in the case of sequences of random variables, just meaning that the relations hold almost surely (and therefore N_0 may depend on ω). For $A \subset I$, (not necessarily in $\mathcal{P}$), we set

$$\gamma(A) \stackrel{\mathrm{def}}{=} \sum_{i\in A} \gamma_i, \ \ \alpha(A) \stackrel{\mathrm{def}}{=} \sum_{J\in\mathcal{P}_A} a_A.$$

We rewrite F_N in terms of energy levels. For a collection $\lambda = (\lambda_J)_{J\in\mathcal{P}}, \ \lambda_J \in \mathbb{R}$, and $A \subset I$, we set

$$\mathcal{N}_{N,A}(\lambda) \stackrel{\mathrm{def}}{=} \#\left\{\sigma \in \Sigma_{N,A} : X_{\sigma_J}^J \geq \lambda_J N, \ \forall J \in \mathcal{P}_A\right\},$$
$$\mathcal{N}_N(\lambda) \stackrel{\mathrm{def}}{=} \mathcal{N}_{N,I}(\lambda).$$

Clearly

$$\{\mathcal{N}_{N,A}(\lambda) = 0\} \subset \{\mathcal{N}_N(\lambda) = 0\}. \tag{1.20}$$

We express F_N in terms of the $\mathcal{N}_N(\lambda)$:

$$\begin{aligned} F_N(\beta) &= \frac{1}{N}\log 2^{-N}(\beta N)^{|\mathcal{P}|}\int_{\mathbb{R}^{\mathcal{P}}} d\lambda \mathcal{N}_N(\lambda)\prod_{J\in\mathcal{P}} \mathrm{e}^{\beta\lambda_J N} \\ &= \frac{1}{N}\log\int_{\mathbb{R}^{\mathcal{P}}} d\lambda\mathcal{N}_N(\lambda)\prod_{J\in\mathcal{P}}\mathrm{e}^{\beta\lambda_J N} - \log 2 + O\left(\frac{\log N}{N}\right). \end{aligned} \tag{1.21}$$

We first want to take out the λ for which $\mathcal{N}_N(\lambda) = 0$ for large N. As these are integer valued random variables, it is clear that $\mathbb{E}\mathcal{N}_N(\lambda) \ll 1$ implies $\mathcal{N}_N(\lambda) = 0$ for large enough N, almost surely. It however turns out, that this condition is not sufficient for our purpose, but remark that if for *some* $A \subset I$, one has $\mathbb{E}\mathcal{N}_{N,A}(\lambda) \ll 1$, then by (1.20), one has $\mathcal{N}_N(\lambda) = 0$ for large enough N as well.

Lemma 1.14. a) *For any $\lambda \in \mathbb{R}^{\mathcal{P}}$ and $A \subset I$ we have*

$$\mathbb{E}\mathcal{N}_{N,A}(\lambda) \asymp 2^{\gamma(A)N}\exp\left[-\sum_{J\in\mathcal{P}_A}\frac{\left(\lambda_J^+\right)^2}{2a_J}N\right],$$

where $\lambda_J^+ \stackrel{\mathrm{def}}{=} \max(\lambda_J, 0)$.

b) *There exists* $C > 0$ *such that*

$$\mathbb{E}\mathcal{N}_N(\lambda) \leq C2^N \exp\left[-\sum_{J\in\mathcal{P}} \frac{\left(\lambda_J^+\right)^2}{2a_J} N\right]$$

for all $\lambda \in \mathbb{R}^{\mathcal{P}}$, *and all* N.

c) *Let* $\lambda \in \mathbb{R}^{\mathcal{P}}$. *If for some* $A \subset I$ *one has*

$$\sum_{J\in\mathcal{P}_A} \frac{\left(\lambda_J^+\right)^2}{2a_J} > \gamma(A)\log 2.$$

then $\mathbb{P}(\mathcal{N}_N(\lambda) \neq 0) \ll 1$, *and in particular* $\mathcal{N}_N(\lambda) = 0$ *for large enough* N, *almost surely.*

Proof. a) and b) follow by standard Gaussian tail estimates, and in case c), by (1.20) we have

$$\mathbb{P}(\mathcal{N}_N(\lambda) \neq 0) \leq \mathbb{P}(\mathcal{N}_{N,A}(\lambda) \neq 0) \leq \mathbb{E}\mathcal{N}_{N,A}(\lambda)$$

which proves c). □

Let

$$\Delta \overset{\text{def}}{=} \left\{\lambda \in \mathbb{R}^{\mathcal{P}} : \sum\nolimits_{J\in\mathcal{P}_A} \frac{\left(\lambda_J^+\right)^2}{2a_J} \leq \gamma(A)\log 2,\ \forall A \subset I\right\},$$

$$\Delta^+ \overset{\text{def}}{=} \{\lambda \in \Delta : \lambda_J \geq 0,\ \forall J \in \mathcal{P}\}.$$

Lemma 1.15. *If* $\lambda \in \operatorname{int}\Delta$ *then*

$$\mathcal{N}_N(\lambda) \asymp \mathbb{E}\mathcal{N}_N(\lambda) \asymp \exp\left[N\left(\log 2 - \sum\nolimits_{J\in\mathcal{P}} \left(\lambda_J^+\right)^2/2a_J\right)\right] \tag{1.22}$$

Proof. The second relation is Lemma 1.14 a). For the proof of the first, it suffices to show that $\lambda \in \operatorname{int}\Delta$ implies

$$\operatorname{var}\mathcal{N}_N(\lambda) \ll (\mathbb{E}\mathcal{N}_N(\lambda))^2. \tag{1.23}$$

In fact, from (1.23), Tchebyshev inequality and the Borel-Cantelli Lemma immediately imply (1.22).

We abbreviate $\mathbb{P}\left(X_{\sigma_J}^J \geq N\lambda_J\right)$ by $p_J(N)$ (λ is kept fixed through this proof). With this notation $\mathbb{E}\mathcal{N}_N(\lambda) = 2^N \prod_{J\in\mathcal{P}} p_J(N)$.

$$\begin{aligned}
\mathbb{E}\mathcal{N}_N(\lambda)^2 &= \sum_{\sigma,\sigma'\in\Sigma_N} \prod_{J\in\mathcal{P}} \mathbb{P}\left(X_{\sigma_J}^J \geq N\lambda_J,\ X_{\sigma'_J}^J \geq N\lambda_J\right) \\
&= \sum_{A\subset I} \sum_{(\sigma,\sigma')\in\Lambda_A} \prod_{J\in\mathcal{P}} \mathbb{P}\left(X_{\sigma_J}^J \geq N\lambda_J,\ X_{\sigma'_J}^J \geq N\lambda_J\right) \\
&= \sum_{A\subset I} |\Lambda_A(N)| \prod_{J\in\mathcal{P}_A} p_J(N) \prod_{J\in\mathcal{P}\setminus\mathcal{P}_A} p_J(N)^2.
\end{aligned}$$

where $\Lambda_A(N)$ consists of those pairs (σ, σ') which agree on A and disagree on $I \backslash A$. For $A = \varnothing$, $2^{2N} - |\Lambda_\varnothing(N)| \ll 2^{2N}$, and therefore

$$\operatorname{var} \mathcal{N}_N(\lambda) = \sum_{A \neq \emptyset} |\Lambda_A(N)| \prod_{J \in \mathcal{P}_A} p_J(N) \prod_{J \in \mathcal{P} \backslash \mathcal{P}_A} p_J(N)^2 + \Omega\left((\mathbb{E}\mathcal{N}_N(\lambda))^2\right).$$

$$|\Lambda_A(N)| = 2^{\gamma(A)N} \prod\nolimits_{i \notin A} 2^{\gamma_i N}\left(2^{\gamma_i N} - 1\right) = 2^{2N} 2^{-\gamma(A)N} + \Omega\left(|\Lambda_A(N)|\right).$$

As by assumption

$$2^{-\gamma(A)N} \ll \prod_{J \in \mathcal{P}_A} p_J(N) \asymp \exp\left[-\sum\nolimits_{J \in \mathcal{P}_A} \frac{\left(\lambda_J^+\right)^2}{2a_J} N\right],\ A \neq \varnothing,$$

we have for any $A \neq \varnothing$

$$|\Lambda_A(N)| \prod_{J \in \mathcal{P}_A} p_J(N) \prod_{J \in \mathcal{P} \backslash \mathcal{P}_A} p_J(N)^2 \ll 2^{2N} \prod_{J \in \mathcal{P}} p_J(N)^2,$$

proving $\operatorname{var} \mathcal{N}_N(\lambda) \ll (\mathbb{E}\mathcal{N}_N(\lambda))^2$. □

Let

$$\psi(\lambda, \beta) \stackrel{\text{def}}{=} \sum_{J \in \mathcal{P}} \left(\beta\lambda_J - \frac{\lambda_J^2}{2a_J}\right). \tag{1.24}$$

Proposition 1.16. *The free energy as defined in (1.5) exists and is given as*

$$f(\beta) = \sup_{\lambda \in \Delta^+} \psi(\lambda, \beta). \tag{1.25}$$

Proof. We show the lower bound for $\liminf_{N\to\infty} F_N(\beta)$, and the upper bound for $\limsup_{N\to\infty} f_N(\beta)$. By the self-averaging property (3.29), this proves the statement. We use the integral representation (1.21). If $\mu = (\mu_J)$, $\nu = (\nu_J)$ satisfy $\mu_J < \nu_J$ for all J, we write

$$[\mu, \nu) \stackrel{\text{def}}{=} \{\lambda : \mu_J \leq \lambda_J < \nu_J,\ \forall J \in \mathcal{P}\}.$$

If $[\mu, \nu) \subset \Delta^+$, we have

$$\begin{aligned} \liminf_{N\to\infty} F_N(\beta) &\geq \liminf_{N\to\infty} \frac{1}{N} \log \int_{[\mu,\nu)} d\lambda \mathcal{N}_N(\lambda) \prod_{J \in \mathcal{P}} e^{\beta\lambda_J N} \\ &\geq \liminf_{N\to\infty} \frac{1}{N} \log \mathcal{N}_N(\nu) \int_{[\mu,\nu)} d\lambda \prod_{J \in \mathcal{P}} e^{\beta\lambda_J N} \\ &\geq \sum\nolimits_J \left(\beta\mu_J - \frac{\nu_J^2}{2a_J}\right). \end{aligned}$$

As this holds for arbitrary $[\mu, \nu) \subset \Delta^+$, $\liminf_{N\to\infty} F_N(\beta) \geq \sup_{\lambda \in \Delta^+} \psi(\lambda, \beta)$ follows.

For the upper bound, let Δ_ε be an ε-neighborhood of Δ. From Lemma 1.14 c), we have

$$\mathbb{P}\left(F_N(\beta) \neq \frac{1}{N} \log \int_{\Delta_\varepsilon} d\lambda \mathcal{N}_N(\lambda) e^{\beta \sum_J \lambda_J N}\right) \ll 1.$$

and therefore

$$\left|\mathbb{E}F_N(\beta) - \frac{1}{N}\mathbb{E}\log\int_{\Delta_\varepsilon} d\lambda\mathcal{N}_N(\lambda)\,\mathrm{e}^{\beta\sum_J \lambda_J N}\right| \ll 1.$$

By Jensen's inequality and Lemma 1.14 b) we have

$$\begin{aligned}
&\limsup_{N\to\infty}\frac{1}{N}\mathbb{E}\log\int_{\Delta_\varepsilon} d\lambda\mathcal{N}_N(\lambda)\,\mathrm{e}^{\beta\sum_J \lambda_J N}\\
&\leq \limsup_{N\to\infty}\frac{1}{N}\log\int_{\Delta_\varepsilon} d\lambda\mathbb{E}\mathcal{N}_N(\lambda)\,\mathrm{e}^{\beta\sum_J \lambda_J N}\\
&\leq \limsup_{N\to\infty}\frac{1}{N}\log 2^N\int_{\Delta_\varepsilon} d\lambda\exp\left[N\sum_{J\in\mathcal{P}}\left(\beta\lambda_J - \frac{\left(\lambda_J^+\right)^2}{2a_J}\right)\right]\\
&\leq \sup_{\lambda\in\Delta_\varepsilon}\psi(\lambda,\beta).
\end{aligned}$$

As $\varepsilon > 0$ is arbitrary, $\limsup_{N\to\infty}\mathbb{E}F_N(\beta) \leq \sup_{\lambda\in\Delta^+}\psi(\lambda,\beta)$ follows. □

1.2.2. The optimization problem. We first discuss the special case of a GREM. Therefore, we assume that the sets in $\mathcal{P}$ are nested, i.e. $\mathcal{P} = \{J_1,\ldots,J_m\}$, where $\emptyset \subset J_1 \subset \ldots \subset J_m$. If $A \subset I$, put $l_A \stackrel{\mathrm{def}}{=} \max\{l : J_l \subset A\}$. Evidently

$$\sum_{J\in\mathcal{P}_A}\frac{\lambda_J^2}{2a_J} \leq \log 2\ \gamma(A)$$

follows from

$$\sum_{i=1}^{l_A}\frac{\lambda_{J_i}^2}{2a_{J_i}} \leq \log 2\ \gamma(J_{l_A}).$$

Therefore, $\lambda \in \Delta^+$ is equivalent with

$$\sum_{i=1}^{l}\frac{\lambda_{J_i}^2}{2a_{J_i}} \leq \log 2\ \gamma(J_l),\ 1 \leq l \leq m,$$

(and of course that all components are non-negative). Therefore we have proved

Lemma 1.17. *Assume that $\mathcal{P}$ is nested as above. Then*

$$f(\beta) = \sup\left\{\psi(\lambda,\beta) : \sum\nolimits_{i=1}^{l}\frac{\lambda_{J_i}^2}{2a_{J_i}} \leq \log 2\ \gamma(J_l),\ 1 \leq l \leq m\right\}.$$

This lemma proves that in our more general situation, for any chain $\mathbf{T}$, the corresponding GREM free energy is an upper bound.

Corollary 1.18. *For any chain $\mathbf{T}$ we have*

$$f(\beta) \leq f(\mathbf{T},\beta),\ \beta \geq 0.$$

Proof. For a given chain $\emptyset = A_0 \subset A_1 \subset A_2 \subset \ldots \subset A_K = I$, we consider $\Delta^+_{\mathbf{T}}$ which is obtained by dropping the conditions for the A's which are not in the chain. Then

$$f(\beta) = \sup_{\lambda\in\Delta^+} \psi(\lambda,\beta) \leq \sup_{\lambda\in\Delta^+_{\mathbf{T}}} \psi(\lambda,\beta),$$

We claim that

$$\sup_{\lambda\in\Delta^+_{\mathbf{T}}} \psi(\lambda,\beta) = f(\mathbf{T},\beta),$$

which proves the corollary. To see this equation, we write

$$\psi(\lambda,\beta) = \sum_{j=1}^{K} \sum_{J\in\mathcal{P}_{A_j}\setminus\mathcal{P}_{A_{j-1}}} \left(\beta\lambda_J - \frac{\lambda_J^2}{2a_J}\right) \stackrel{\text{def}}{=} \sum_{j=1}^{K} \psi_j(\lambda_j,\beta), \text{ say,}$$

where $\lambda_j \stackrel{\text{def}}{=} (\lambda_J)_{J\in\mathcal{P}_{A_j}\setminus\mathcal{P}_{A_{j-1}}}$. Set

$$f_j(\beta,t) \stackrel{\text{def}}{=} \sup\left\{\psi_j(\lambda_j,\beta) : \sum_{J\in\mathcal{P}_{A_j}\setminus\mathcal{P}_{A_{j-1}}} \frac{\lambda_J^2}{2a_J} = t\right\} = \beta\sqrt{2t\hat{a}_j} - t,$$

where $\hat{a}_j \stackrel{\text{def}}{=} \sum_{J\in\mathcal{P}_{A_j}\setminus\mathcal{P}_{A_{j-1}}} a_J$, i.e. $f_j\left(\beta, s^2/2\hat{a}_j\right) = \beta s - s^2/2\hat{a}_j$. We therefore see that

$$\begin{aligned}\sup_{\lambda\in\Delta^+_{\mathbf{T}}} \psi(\lambda,\beta) &= \sup\left\{\sum\nolimits_{j=1}^{K}\left(\beta s_j - \frac{s_j^2}{2\hat{a}_j}\right) : \sum\nolimits_{j=1}^{l} \frac{s_j^2}{2\hat{a}_j} \leq \log 2\ \gamma(A_l),\ 1\leq l\leq K\right\}\\ &= f(\mathbf{T},\beta),\end{aligned}$$

the last equality by Lemma 1.17. $\square$

In order to finish the proof of Theorem 1.1, it only remains to construct a chain $\mathbf{T}$ which satisfies $f(\beta) \geq f(\mathbf{T},\beta)$. (Then one has also equality by the above corollary.) This is done by the algorithmic prescription described before of Proposition 1.2: first remark that the conditions **C1(k)** and **C2(k)** are void for $k=0$. If $A_k = I$, then the construction is finished, and we have $K \stackrel{\text{def}}{=} k$. Therefore, assume $A_k \neq I$. Then we set $\beta_{k+1} \stackrel{\text{def}}{=} \hat{\rho}(A_k)$, and prove first that $\beta_{k+1} > \beta_k$. We claim that for any $A \supset A_k$, $A \neq A_k$, one has

$$2\log 2\left(\gamma(A) - \gamma(A_k)\right) > \beta_k^2\left(\alpha(A) - \alpha(A_k)\right).$$

Indeed, because of $2\log 2\left(\gamma(A_k) - \gamma(A_{k-1})\right) = \beta_k^2\left(\alpha(A_k) - \alpha(A_{k-1})\right)$,

$$2\log 2\left(\gamma(A) - \gamma(A_k)\right) < \beta_k^2\left(\alpha(A) - \alpha(A_k)\right)$$

would contradict Condition **C1(k)** and equality would contradict **C2(k)**.

It only remains to construct A_{k+1} which satisfies **C2(k+1)**. Assume there are two sets $A, A' \supset A_k$, $A, A' \neq A_k$ satisfying

$$\rho(A_k, A) = \rho\left(A_k, A'\right) = \beta_{k+1}. \tag{1.26}$$

We claim that then also $\rho(A_k, A \cup A') = \beta_{k+1}$. Remark that

$$\alpha\left(A \cup A'\right) \geq \alpha(A) + \alpha\left(A'\right) - \alpha\left(A \cap A'\right),$$
$$\gamma\left(A \cup A'\right) = \gamma(A) + \gamma\left(A'\right) - \gamma\left(A \cap A'\right),$$

and therefore

$$\begin{aligned}
&2\log 2\left(\gamma\left(A \cup A'\right) - \gamma(A_k)\right) - \beta_{k+1}^2\left(\alpha\left(A \cup A'\right) - \alpha(A_k)\right) \\
&\leq 2\log 2\left[\gamma(A) + \gamma\left(A'\right) - \gamma\left(A \cap A'\right) - \gamma(A_k)\right] \\
&\quad - \beta_{k+1}^2\left[\alpha(A) + \alpha\left(A'\right) - \alpha\left(A \cap A'\right) - \alpha(A_k)\right] \\
&= \beta_{k+1}^2\left[\alpha\left(A \cap A'\right) - \alpha(A_k)\right] - 2\log 2\left[\gamma\left(A \cap A'\right) - \gamma(A_k)\right] \leq 0,
\end{aligned}$$

the equality by (1.26), and the last inequality by the definition of β_{k+1}. From the definition of β_{k+1}, we therefore conclude that

$$2\log 2\left(\gamma\left(A \cup A'\right) - \gamma(A_k)\right) = \beta_{k+1}^2\left(\alpha\left(A \cup A'\right) - \alpha(A_k)\right)$$

We therefore find a *unique* maximal set $A_{k+1} \supset A_k$ which satisfies $\rho(A_k, A_{k+1}) = \beta_{k+1}$, and so we have constructed $\beta_{k+1} > \beta_k$, $A_{k+1} \supset A_k$, $A_{k+1} \neq A_k$ such that **C1(k+1)** and **C2(k+1)** are satisfied. The construction terminates after a finite number of steps.

We claim now that with $\mathbf{T} \stackrel{\text{def}}{=} (\emptyset, A_1, \ldots, A_{K-1}, I)$ we have

$$f(\beta) \geq f(\beta, \mathbf{T}). \tag{1.27}$$

Clearly, if $\beta > 0$ is small enough, the maximum in (1.25) is attained in $\lambda_J^{(1)}(\beta) \stackrel{\text{def}}{=} a_J\beta$ for all J, and therefore

$$f(\beta) = \frac{\beta^2}{2}$$

for small β. This remains valid as long as $\left(\beta^2/2\right)\alpha(A) \leq \gamma(A)$ for all A, i.e. for $\beta \leq \beta_1$. For $\beta_k < \beta \leq \beta_{k+1}$, we choose $\lambda^{(k+1)}(\beta)$ defined by

$$\lambda_J^{(k+1)}(\beta) \stackrel{\text{def}}{=} \begin{cases} a_J\beta_m & \text{for } J \in \mathcal{P}_{A_m} \setminus \mathcal{P}_{A_{m-1}},\ 1 \leq m \leq k \\ a_J\beta & \text{for } J \notin \mathcal{P}_{A_k} \end{cases}. \tag{1.28}$$

This choice (1.28) satisfies the side conditions in the range of β we are considering, and hence

$$\psi\left(\lambda^{(k+1)}(\beta), \beta\right) \leq f(\beta). \tag{1.29}$$

We show now that $f(\beta, \mathbf{T}) = \psi\left(\lambda^{(k+1)}(\beta), \beta\right)$ for $\beta_k \leq \beta \leq \beta_{k+1}$. An elementary computation gives

$$\begin{aligned}
\psi\left(\lambda^{(k+1)}(\beta), \beta\right) &= \beta\sum_{i=1}^{k}\beta_i\left[\alpha(A_i) - \alpha(A_{i-1})\right] - \gamma(A_k)\log 2 + \frac{\beta^2}{2}(1 - \alpha(A_k)) \\
&= \beta\sum_{i=1}^{k}\beta_i\hat{a}(A_i) - \gamma(A_k)\log 2 + \frac{\beta^2}{2}\sum_{i=k+1}^{K}\hat{a}(A_i),
\end{aligned}$$

where $\hat{a}(A_i)$ is defined by (1.4). This is exactly the free energy of the corresponding GREM as given in [**11**]. (It is in fact elementary to check that $\lambda^{(k+1)}(\beta)$ is the maximizing vector λ for the GREM corresponding to the above chain when $\beta_k \leq \beta \leq \beta_{k+1}$.)

We have therefore proved Theorem 1.1 as well as the representation given in Proposition 1.2.

□

1.3. Gibbs measure of the non hierarchical GREM, Proofs

1.3.1. Localization of the energy levels. The goal of this section is to prove Proposition 1.5. Let us begin with some fairly evident estimates:

$$\frac{a_{N,j}}{\Delta_j N} = \beta_j + O(N^{-1}\log N), \quad \exp\Big[-\frac{a_{N,j}^2}{2\Delta_j N}\Big] = 2^{-G_j N}\beta_j\sqrt{2\pi\Delta_j N}\big[1+o(1)\big]. \tag{1.30}$$

The next Lemma relates to exponentials of gaussian random variables. Let $B > \beta_j$ and $B_N \stackrel{\text{def}}{=} B + \epsilon_N$, for some $\epsilon_N \to 0$.

Lemma 1.19. *For any sequence of reals $\phi_1, \ldots, \phi_j$ there exists "const" depending on the underlying parameters only (not yet on $\phi's$) such that for N large enough*

$$\mathbb{E}\Big[\exp\Big(B_N\widehat{Y}_j\Big);\ \widehat{Y}_1 \le \phi_1, \widehat{Y}_2 \le \phi_2, \ldots, \widehat{Y}_j \le \phi_j\Big]$$
$$\lesssim 2^{-\gamma(A_j)N}\exp\Big\{\sum_{l=1}^{j-1}(\beta_{l+1}-\beta_l)\phi_l + (B-\beta_j)\phi_j\Big\}. \tag{1.31}$$

Proof. Let $\mathbb{E}_{\overline{Y}_j}$ stand for expectation w.r.t. $\overline{Y}_j$. Then

$$\mathbb{E}\Big[\exp\Big(B_N\widehat{Y}_j\Big);\ \widehat{Y}_1 \le \phi_1, \widehat{Y}_2 \le \phi_2, \ldots, \widehat{Y}_j \le \phi_j\Big] =$$
$$= \mathbb{E}\Big[\exp\Big(B_N\widehat{Y}_{j-1}\Big)\mathbb{E}_{\overline{Y}_j}\Big[\exp\big(B_N\overline{Y}_j\big);\ \widehat{Y}_{j-1}+\overline{Y}_j \le \phi_j\Big];\ \widehat{Y}_1 \le \phi_1, \ldots, \widehat{Y}_{j-1} \le \phi_{j-1}\Big]. \tag{1.32}$$

But

$$\mathbb{E}_{\overline{Y}_j}\Big[\exp\big(B_N\overline{Y}_j\big);\ \widehat{Y}_{j-1}+\overline{Y}_j \le \phi_j\Big] = \int_{-\infty}^{\phi_j-\widehat{Y}_{j-1}}\exp\Big[B_N x - \frac{\big(x+a_{N,j}\big)^2}{2\Delta_j N}\Big]\frac{dx}{\sqrt{2\pi\Delta_j N}}$$
$$\le \exp\Big[-\frac{a_N,j^2}{2\Delta_j N}\Big]\times\int_{-\infty}^{\phi_j-\widehat{Y}_{j-1}}\exp\Big[\Big(B_N-\frac{a_{N,j}}{\Delta_j N}\Big)x\Big]\frac{dx}{\sqrt{2\pi\Delta_j N}}. \tag{1.33}$$

Observe that, for N large enough, $B_N - \frac{a_{N,j}}{N\Delta_j}$ is strictly positive (it converges to $B-\beta_j$), whence the existence of the last integral above, which together with the bounds (1.30) leads to

$$(1.33) \lesssim 2^{-G_j N}\exp\Big[\Big(B_N - \frac{a_{N,j}}{\Delta_j N}\Big)(\phi_j - \widehat{Y}_{j-1})\Big]. \tag{1.34}$$

Plugging (1.34) into (1.32) and iterating the procedure with B_N replaced by $\frac{a_{N,j}}{N\Delta_j} = \beta_j + \tilde{\epsilon}_N$ (with some new $\tilde{\epsilon}_N \to 0$) yields the claim. $\square$

A first application of the above is:

Lemma 1.20. *To bounded $\mathfrak{M} \subset \mathbb{R}$, the following hold for N large enough:*

a) *There exists C such that* $\mathbb{P}\Big[\exists \tau \in \Sigma_{N,A_j} : \widehat{X}_{\tau(1),\dots,\tau(l)} \geq C$ *for some* $l \leq j\Big] \leq \epsilon.$

b) *There exists $\widehat{R} > 0$ such that*

$$\mathbb{P}\Big[\exists \tau \in \Sigma_{N,A_j} : \widehat{X}_{\tau(1),\dots,\tau(j)} \in \mathfrak{M},\ \widehat{X}_{\tau(1),\dots,\tau(l)} \notin (-\widehat{R}, \widehat{R})\, \textit{for some}\ l \leq j\Big] \leq \epsilon.$$

Proof. The first claim goes best by induction on the level j. Suppose that there exists $\widehat{C}$ such that

$$\mathbb{P}\Big[\forall \tau \in \Sigma_{N,A_l} :\ \widehat{X}_{\tau(1),\dots,\tau(l)} \leq \widehat{C},\ \forall\, l \leq j-1\Big] \geq 1 - \epsilon/2$$

for N large enough. For any $\widetilde{C} > 0$ we thus have

$$\begin{aligned} \mathbb{P}\Big[\exists \tau \in \Sigma_{N,A_j} : \widehat{X}_{\tau(1),\dots,\tau(j)} \geq \widetilde{C}\Big] &\leq \frac{\epsilon}{2} + \\ &+ \mathbb{P}\Big[\exists \tau \in \Sigma_{N,A_j} :\ \widehat{X}_{\sigma(1),\dots,\sigma(j)} \geq \widetilde{C} \text{ and } \forall l \leq (j-1)\ \ \widehat{X}_{\tau(1),\dots,\tau(l)} \leq \widehat{C}\Big], \end{aligned} \tag{1.35}$$

and the second term on the r.h.s above is bounded by

$$\begin{aligned} &\sum_{\tau \in \Sigma_{N,A_j}} \mathbb{P}\Big[\widehat{X}_{\tau(1)} \leq \widehat{C}, \dots, \widehat{X}_{\tau(1),\dots,\tau(j-1)} \leq \widehat{C}, \widehat{X}_{\tau(1),\dots,\tau(j)} \geq \widetilde{C}\Big] \\ &= 2^{\gamma(A_j)N} \mathbb{P}\Big[\widehat{Y}_1 \leq \widehat{C}, \dots, \widehat{Y}_{j-1} \leq \widehat{C}, \overline{Y}_j \geq \widetilde{C} - \widehat{Y}_{j-1}\Big] \\ &= 2^{\gamma(A_j)N} \mathbb{E}\left[\int_{\widetilde{C} - \widehat{Y}_{j-1}}^{\infty} \exp\Big[-\frac{(x + a_{N,j})^2}{2\Delta_j N}\Big] \frac{dx}{\sqrt{2\pi\Delta_j N}};\ \widehat{Y}_1 \leq \widehat{C}, \dots, \widehat{Y}_{j-1} \leq \widetilde{C}\right] \\ &\lesssim 2^{\gamma(A_j)N} \mathbb{E}\left[\exp\Big[-\frac{a_{N,j}^2}{\Delta_j N} - \frac{a_{N,j}}{2\Delta_j N}\big(\widetilde{C} - \widehat{Y}_{j-1}\big) + o(1)\Big]; \widehat{Y}_1 \leq \widehat{C}, \dots, \widehat{Y}_{j-1} \leq \widetilde{C}\right] \\ &\overset{\text{Lemma 1.19}}{\lesssim} \exp\left[\sum_{l=1}^{j-1} (\beta_{l+1} - \beta_l)\widehat{C} - \beta_j \widetilde{C}\right]. \end{aligned} \tag{1.36}$$

It thus suffices to choose $\widetilde{C}$ large enough in the positive to make the above less then $\epsilon/2$. Setting $C \stackrel{\text{def}}{=} \max\{\widetilde{C}, \widehat{C}\}$ yields a). As for the proof of claim b), let $\widetilde{C} > 0$ and $\overline{\mathfrak{M}} \stackrel{\text{def}}{=} \sup\{x \in \mathfrak{M}\}$. By a) we can find $C > 0$ such that for large enough N

$$\mathbb{P}\big[\forall \tau \in \Sigma_{N,A_j} : \widehat{X}_{\tau(1),\dots,\tau(j)} \leq C \,\text{for all}\, l \leq j\big] \geq 1 - \epsilon/4. \tag{1.37}$$

and therefore

$$\begin{aligned}\mathbb{P}\Big[\exists\tau\in\Sigma_{N,A_j}\,:\;\widehat{X}_{\tau(1),\dots,\tau(j)}\in\mathfrak{M},\;\widehat{X}_{\tau(1),\dots,\tau(l)}\leq-\widetilde{C}\text{ for some } l\leq j\Big]&\\ \leq\epsilon/4+\mathbb{P}\Big[\exists\tau\in\Sigma_{N,A_j}\,:\;\widehat{X}_{\tau(1),\dots,\tau(j)}\in\mathfrak{M},\;\widehat{X}_{\tau(1),\dots,\tau(l)}\leq-\widetilde{C}&\\ \text{for some } l\leq j,\quad \widehat{X}_{\tau(1),\dots,\tau(r)}\leq C\;\forall r\leq j\Big]&\\ \leq\epsilon/4+const\times\sum_{l\leq j}\exp\left[\sum_{k\neq l}(\beta_{k+1}-\beta_k)\max(C,\overline{\mathfrak{M}})-(\beta_{l+1}-\beta_l)\widetilde{C}\right].&\end{aligned} \tag{1.38}$$

(the steps behind the last inequality following verbatim those in (1.36)). It thus suffices to choose $\widetilde{C}$ large enough in the positive to make (1.38) smaller then say $\frac{3}{4}\epsilon$, which together with (1.37) yields the claim of part b) with $\widehat{R}=\max(C,\widetilde{C})$.

□

Proof of Proposition 1.5. The claim follows steadily from part b) of Lemma 1.20 since within our notational convention $\overline{X}_{\tau(1),\dots,\tau(l)}=\widehat{X}_{\tau(1),\dots,\tau(l)}-\widehat{X}_{\tau(1),\dots,\tau(l-1)}$. □

1.3.2. Random configuration spaces. The goal of this section is to prove Proposition 1.6, for which we need some additional results pertaining to the asymptotics of "rare events". For bounded $\mathfrak{M}\subset\mathbb{R}$, we set $p_N(j,\mathfrak{M})\stackrel{\text{def}}{=}\mathbb{P}\Big[\overline{Y}_j\in\mathfrak{M}\Big]$. Let $\varepsilon>0$ and $\eta\in(0,1/2)$. For critical $A\subsetneq A_j\setminus A_{j-1}$ we write

$$p_N(j,\mathfrak{M},A;\varepsilon,\eta)\stackrel{\text{def}}{=}\mathbb{P}\Big[\overline{Y}_j\in\mathfrak{M},\frac{Y_{j,A}}{\widehat{\alpha}_j(A)}-\frac{Y^c_{j,A}}{\widehat{\alpha}^c_j(A)}\geq-\varepsilon,\sqrt{N}Y_{j,A}-a_{N,j}(A)\leq N^\eta\Big],$$

For non-critical $A\subsetneq A_j\setminus A_{j-1}$ such that $\widehat{\alpha}_j(A)>0$,

$$p_N^>(\mathfrak{M},j,A,\varepsilon)\stackrel{\text{def}}{=}\mathbb{P}\Big[\overline{Y}_j\in\mathfrak{M},Y_{j,A}>\beta_j\widehat{\alpha}_j(A)(1+\varepsilon)\sqrt{N}\Big]$$

Lemma 1.21. *For N large enough:*

a) $p_N(j,\mathfrak{M})=2^{-G_jN}\int_{\mathfrak{M}}\beta_j\exp\big[-\beta_jx+o(1)\big]dx$,
b) $p_N^>(\mathfrak{M},j,A,\varepsilon)\lesssim2^{-G_jN}\exp\big[-const\times\varepsilon^2N\big]$.
c) $p_N(j,\mathfrak{M},A;\varepsilon,\eta)\lesssim2^{-G_jN}\times\varepsilon$.

Proof. Claim $a)$ and $b)$ easily follow from the asymptotics (1.30). To prove c), first recall that $a_{N,j} = a_{N,j}(A) + \beta_j \widehat{\alpha}_j^c(A)N + O(1)$ and therefore

$$p_N(j, \mathfrak{M}, A; \varepsilon, \eta) \lesssim \frac{1}{\sqrt{N}} \int_{-\infty}^{N^\eta} \exp\Big[-\big(x + a_{N,j}(A)\big)^2 / 2\widehat{\alpha}_j(A)N \Big] \frac{dx}{\sqrt{2\pi\widehat{\alpha}_j(A)N}} \times$$
$$\times \int_{\mathfrak{N}_x} \exp\Big[-\big(y + \beta_j \widehat{\alpha}_j^c(A)N\big)^2 / 2\widehat{\alpha}_j^c(A)N \Big] dy,$$

with $\mathfrak{N}_x \stackrel{\text{def}}{=} \Big\{\mathfrak{M} - x + O(1)\Big\} \cap \Big\{ y \in \mathbb{R} : \dfrac{x}{\sqrt{N}\widehat{\alpha}_j(A)} - \dfrac{y}{\sqrt{N}\widehat{\alpha}_j^c(A)} \geq -\varepsilon + O(\log N/\sqrt{N}) \Big\}$. (1.39)

Since $\mathfrak{M}$ is bounded, for the integration set $\mathfrak{N}_x$ not to be empty we must have $x \geq x_{min} \stackrel{\text{def}}{=} -const \cdot \varepsilon \cdot \sqrt{N} + O(\log N)$, with $const = \widehat{\alpha}_j(A)\widehat{\alpha}_j^c(A)/\Delta_j$. Therefore:

$$\begin{aligned}
(1.39) &\lesssim \frac{1}{\sqrt{N}} \exp\Big[-\frac{\beta_j^2}{2} \widehat{\alpha}_j^c(A)N \Big] \int_{\mathfrak{M}} \exp\big[-\beta_j y\big] dy \times \\
&\qquad \times \int_{x_{min}}^{N^\eta} \exp\big(\beta_j x\big) \exp\left[-\frac{\big(x + a_{N,j}(A)\big)^2}{2\widehat{\alpha}_j(A)N} \right] \frac{dx}{\sqrt{2\pi\widehat{\alpha}_j(A)N}} \\
&\lesssim \frac{1}{\sqrt{N}} \exp\Big[-\frac{\beta_j^2}{2}\widehat{\alpha}_j^c(A)N + \frac{\beta_j^2}{2}\widehat{\alpha}_j(A)N - a_{N,j}(A)\beta_j \Big] \times \\
&\qquad \times \int_{x_{min}}^{N^\eta} \exp\left[-\frac{\big(x + a_{N,j}(A) - \beta_j\widehat{\alpha}_j(A)N\big)^2}{2\widehat{\alpha}_j(A)N} \right] \frac{dx}{\sqrt{2\pi\widehat{\alpha}_j(A)N}} \\
&\lesssim 2^{-G_j N} \times \mathbb{P}\Big[Y_{j,A} \in \Big(x_{min}N^{-1/2}, N^{\eta - 1/2}\Big) + O(\log N/\sqrt{N}) \Big]
\end{aligned} \tag{1.40}$$

the last step by simply noting that $a_{N,j}(A) - \beta_j \widehat{\alpha}_j(A)N = O(\log N)$. Remark that

$$\lim_{N\to\infty} \mathbb{P}\left[Y_{j,A} \in \Big(x_{min}N^{-1/2}, N^{\eta-1/2}\Big) + O(\log N/\sqrt{N}) \right] = \int_{-const\cdot\varepsilon}^{0} \exp\big(-\frac{x^2}{2}\big) \frac{dx}{\sqrt{2\pi}} \lesssim \varepsilon.$$

This settles claim c). □

Proof of Proposition 1.6. Since R is fixed throughout the proof, we abbreviate $\mathfrak{R} \stackrel{\text{def}}{=} (-R, R)$. We bound

$$\mathbb{P}\Big[\Sigma_{N,A_j}^R \setminus \Sigma_{N,A_j}^{R,\varepsilon_1,\varepsilon_2} \neq \emptyset \Big] \leq (I) + (II) + (III),$$

$$(I) \stackrel{\text{def}}{=} \mathbb{P}\left[\exists \sigma \in \Sigma_{N,A_j}^R : \sum_{J \in \widehat{\mathcal{P}}_{A,k}} X_{\sigma_J}^J - a_{N,k}(A) \geq N^\eta \right.$$
$$\left. \text{for some critical } A \subsetneq A_k \setminus A_{k-1}, k = 1, \ldots, j \right],$$

$$(II) \stackrel{\text{def}}{=} \mathbb{P}\Bigg[\exists \sigma \in \Sigma^R_{N,A_j},\ \mathbf{T}_1(\sigma,k,A,\varepsilon_1) \text{ does not hold for critical } A \subsetneq A_k \setminus A_{k-1}$$
$$\text{for some } k=1,\dots,j \text{ but } \sum_{J\in\widehat{\mathcal{P}}_{A,k}} X^J_{\sigma_J} - a_{N,k}(A) \le N^\eta\Bigg],$$

and finally

$$(III) \stackrel{\text{def}}{=} \mathbb{P}\Bigg[\exists \sigma \in \Sigma^R_{N,A_j} \text{ such that } \mathbf{T_2}(\sigma,k,A,\varepsilon_2) \text{ does not hold for some}$$
$$\text{for some } A \subsetneq A_k \setminus A_{k-1}, k=1,\dots,j\Bigg].$$

We provide upper-bounds to $(I),(II)$ and (III).

$$(I) \le \sum_{k=1}^{j} \sum_{\substack{A\subsetneq A_k\setminus A_{k-1}\\ A \text{ critical}}} \mathbb{P}\Bigg[\exists \sigma \in \Sigma_{N,A_{k-1}\cup A}, \text{ such that}$$
$$\forall l=1,\dots,k-1\ \overline{X}_{\sigma(1),\dots,\sigma(l)} \in \mathfrak{R}, \sum_{J\in\widehat{\mathcal{P}}_{N,A_k}} X^{\{J\}}_{\sigma_J} - a_{N,k}(A) \ge N^\eta\Bigg]$$
$$\le \sum_{k=1}^{j} \sum_{\substack{A\subsetneq A_k\setminus A_{k-1}\\ A \text{ critical}}} 2^{\gamma(A_{k-1})N} \left\{\prod_{l=1,\dots,k-1} p_N(l,\mathfrak{R})\right\} 2^{\gamma(A)N}\mathbb{P}\Big[\sqrt{N}Y_{k,A} - a_{N,k}(A) \ge N^\eta\Big] \tag{1.41}$$

It is easily seen that $\mathbb{P}\Big[\sqrt{N}Y_{k,A} - a_{N,k}(A) \ge N^\eta\Big] \lesssim \exp\Big[-\frac{\beta_k^2}{2}\widehat{\alpha}_k(A)N - \mathit{const}\times N^\eta\Big]$ for some positive *const*, and for critical $A \subsetneq A_k \setminus A_{k-1}$, $\frac{\beta_k^2}{2}\widehat{\alpha}_k(A) = \gamma(A)\log 2$, so it follows from Lemma 1.21 that $(I) \lesssim \exp\big[-\mathit{const}\times N^\eta\big]$ for large enough N.

$$(II) \le \sum_{\sigma\in\Sigma_{N,A_j}} \sum_{\substack{k=1\dots,j\\ A\subsetneq A_k\setminus A_{k-1} \text{ critical}}} \mathbb{P}\Bigg[\overline{X}_{\sigma(1),\dots,\sigma(l)} \in \mathfrak{R}, l \le k,\ \mathbf{T_1}(\sigma,k,A,\varepsilon_1) \text{ holds},$$
$$\sum_{J\in\widehat{\mathcal{P}}_{A,k}} X^{\{J\}}_{\sigma_J} - a_{N,k}(A) \le N^\eta\Bigg]$$
$$\le 2^{\gamma(A_j)N} \sum_{\substack{k=1,\dots,j\\ A\subsetneq A_k\setminus A_{k-1} \text{ critical}}} p_N(k,\mathfrak{R};\varepsilon_1,\eta) \times \prod_{\substack{l=1,\dots,j\\ l\ne k}} p_N(\mathfrak{R},l) \tag{1.42}$$

Hence, by Lemma 1.21, we have $(II) \lesssim \varepsilon_1$ for large enough N. Finally,

$$(III) \leq \sum_{\substack{k=1,\dots,j \\ A\subset A_k \setminus A_{k-1}}} 2^{\gamma(A_k)N} p_N^{>}(\mathfrak{R}, k, A, \varepsilon_2) \prod_{l=1\dots,k-1} p_N(l, \mathfrak{R}) \tag{1.43}$$

which by Lemma 1.21 is easily seen to be $\lesssim \exp[-\mathit{const} \times \varepsilon_2^2 \times N]$ for some positive $\mathit{const} > 0$. Putting the pieces together, we see that $\mathbb{P}\Big[\Sigma^R_{N,A_j} \setminus \Sigma^{R,\varepsilon_1,\varepsilon_2}_{N,A_j}\Big]$ is essentially of order ε_1. □

1.3.3. Suppression of structures and propagation. We first derive some bounds on "two-points probabilities". Although quite straightforward, these estimates are truely essential to rule out non ultrametric configurations. Introduce

$$p_N^{(2)}(j, \mathfrak{M}, A, \varepsilon) \stackrel{\text{def}}{=} \mathbb{P}\Big[\sqrt{N}Y_{j,A} + \sqrt{N}Y^c_{j,A} - a_{N,j} \in \mathfrak{M},$$
$$\sqrt{N}Y_{j,A} + \sqrt{N}Z^c_{j,A} - a_{N,j} \in \mathfrak{M}, Y_{j,A} \leq \beta_j \widehat{\alpha}_j(A)(1+\varepsilon)\sqrt{N}\Big].$$

At last, for critical $A \subsetneq A_j \setminus A_{j-1}$ we write

$$p_N^{(2,crit)}(j, \mathfrak{M}, A, \varepsilon) \stackrel{\text{def}}{=} \mathbb{P}\Bigg[\sqrt{N}Y_{j,A} + \sqrt{N}Y^c_{j,A} - a_{N,j} \text{ and } \sqrt{N}Y_{j,A} + \sqrt{N}Z^c_{j,A} - a_{N,j} \in \mathfrak{M},$$
$$\text{and } \frac{Y_{j,A}}{\widehat{\alpha}_j(A)} - \frac{Y^c_{j,A}}{\widehat{\alpha}^c_j(A)} \leq -\varepsilon,\ \frac{Y_{j,A}}{\widehat{\alpha}_j(A)} - \frac{Z^c_{j,A}}{\widehat{\alpha}^c_j(A)} \leq -\varepsilon\Bigg]$$

Lemma 1.22. *Let $\varepsilon > 0$. For N large enough*

a) $p_N^{(2)}(j, \mathfrak{M}, A, \varepsilon) \lesssim 2^{-2G_jN} \exp\Big\{\beta_j^2 \widehat{\alpha}_j(A)\big[1 - \frac{1}{2}(1-\varepsilon)^2\big]N\Big\}$.
b) $p_N^{(2,crit)}(j, \mathfrak{M}, A, \varepsilon) \lesssim 2^{-2G_jN + \gamma(A)N} \exp\big[-\mathit{const} \times \varepsilon\sqrt{N}\big]$.

Proof. *a)* is straightforward. *b)* Setting $\omega_N = O(\log N)$ for $N \uparrow \infty$, it holds:

$$p_N^{(2,crit)}(j, \mathfrak{M}, A, \varepsilon) \lesssim \int_{-\infty}^{\infty} \exp\left[-\frac{(x + a_{N,j}(A))^2}{2\widehat{\alpha}_j(A)N}\right] dx \left(\int_{\mathfrak{N}} \exp\left[-\frac{\Big(y + \beta_j\widehat{\alpha}^c_j(A)N\Big)^2}{2\widehat{\alpha}^c_j(A)N}\right] dy\right)^2,$$
$$\text{where} \quad \mathfrak{N}_x = \Big\{\mathfrak{M} - x - \omega_N\Big\} \bigcap \Big\{y \in \mathbb{R} :\ y \geq \frac{\widehat{\alpha}^c_j(A)}{\widehat{\alpha}_j(A)}x + \varepsilon\widehat{\alpha}^c_j(A)\sqrt{N} + \omega_N\Big\}. \tag{1.44}$$

$\mathfrak{N}_x$ is not empty as soon as $x \leq x_{\max} \stackrel{\text{def}}{=} -\varepsilon \frac{\widehat{\alpha}_j(A)\widehat{\alpha}_j^c(A)}{\Delta_j}\sqrt{N} + \omega_N$. Thus,

$$
\begin{aligned}
(1.44) \lesssim & \int_{-\infty}^{x_{\max}} \exp\left[-\frac{(x+a_{N,j}(A))^2}{2\widehat{\alpha}_j(A)N}\right] dx \left(\int_{\mathfrak{M}-x-\omega_N} \exp\left[-\frac{\left(y+\beta_j\widehat{\alpha}_j^c(A)\right)^2}{2\widehat{\alpha}_j^c(A)N}\right] dy\right)^2 \\
& \lesssim \exp\Big[-\beta_j^2\widehat{\alpha}_j^c(A)N + \omega_N\Big] \int_{-\infty}^{x_{\max}} \exp\left[-\frac{\left(x-\beta_j\widehat{\alpha}_j(A)N+\omega_N\right)^2}{2\widehat{\alpha}_j(A)N}\right] dx \\
& \lesssim \exp\Big[-\beta_j^2\widehat{\alpha}_j^c(A)N - \frac{\beta_j^2}{2}\widehat{\alpha}_j(A)N + \omega_N\Big] \underbrace{\int_{-\infty}^{x_{\max}} \exp\big[\beta_j x\big] dx}_{\leq \exp\left(-const\times\varepsilon\sqrt{N}\right)} .
\end{aligned} \tag{1.45}
$$

By criticality of A, $\frac{\beta_j^2}{2}\widehat{\alpha}_j(A) = \gamma(A)\log 2$, $\beta_j^2\widehat{\alpha}_j^c(A) = 2\log 2\big[\gamma(A_j) - \gamma(A\cup A_{j-1})\big]$, cfr. Remark 1.10, and therefore

$$
(1.45) \leq 2^{-2G_jN}\exp\big[\gamma(A)N\log 2\big]\exp\big[-const\times\varepsilon\sqrt{N}\big].
$$

□

In the next Proposition we put on rigorous ground the fact that structures in the inner of a given tree-level $(A_j \setminus A_{j-1})$ are suppressed.

Proposition 1.23 (Suppression). *Let σ', τ' be two reference configurations in $\Sigma_{N,A_{j-1}}$. For positive ε_1 and sufficiently small ε_2 there exists const > 0 such that*

$$
\begin{aligned}
\mathbb{P}\Big[\exists\, \sigma,\tau \in \Sigma_{N,A_j}^{R,\varepsilon_1,\varepsilon_2},\ \sigma(j)\neq\tau(j),\ \sigma_{A_{j-1}} = \sigma', \tau_{A_{j-1}} = \tau' : \\
\sigma_s = \tau_s \text{ for some } s\in A_j\setminus A_{j-1}\Big] \lesssim \exp\Big[-const\times\varepsilon_1\sqrt{N}\Big].
\end{aligned} \tag{1.46}
$$

Proof. The l.h.s of (1.46) is clearly bounded by

$$
\begin{aligned}
&\sum_{\substack{A\subsetneq A_j\setminus A_{j-1}\\ A \text{ critical}}} \sum^{\star} \mathbb{P}\Big[\overline{X}_{\sigma(1),\dots,\sigma(j)} \text{ and } \overline{X}_{\tau(1),\dots,\tau(j)} \in \mathfrak{R},\ \mathbf{T}_1(\sigma,j,A,\varepsilon_1),\ \mathbf{T}_1(\tau,j,A,\varepsilon_1) \text{ hold}\Big] + \\
&+ \sum_{\substack{A\subset A_j\setminus A_{j-1}\\ A \text{ non-critical}}} \sum^{\star} \mathbb{P}\Big[\overline{X}_{\sigma(1),\dots,\sigma(j)} \text{ and } \overline{X}_{\tau(1),\dots,\tau(j)} \in \mathfrak{R};\ \mathbf{T}_2(\sigma,j,A,\varepsilon_2), \text{ and } \mathbf{T}_2(\tau,j,A,\varepsilon_2) \text{ hold}\Big].
\end{aligned} \tag{1.47}
$$

In both cases, $\sum^{\star}$ runs over all the $\sigma,\tau\in\Sigma_{N,A_j}$ such that $\sigma(j)\neq\tau(j)$, as well as $\sigma_{A_{j-1}} = \sigma', \tau_{A_{j-1}} = \tau'$, $\sigma_J = \tau_J$ for every $J\in\widehat{\mathcal{P}}_{A,j}$ and $\sigma_J\neq\tau_J$ for every $J\in\widehat{\mathcal{P}}_{A,j}^c$. To fixed $A\subset A_j\setminus A_{j-1}$ there are at most $2^{2G_jN}2^{-\gamma(A)N}$ couples of σ,τ satisfying these requirements. Thus we may upper bound (1.47) by

$$
\begin{aligned}
&\sum_{\substack{A\subsetneq A_j\setminus A_{j-1}\\ A \text{ critical}}} 2^{2G_jN}2^{-\gamma(A)N}p_N^{(2,crit)}(j,\mathfrak{R},A,\varepsilon_1)+\sum_{\substack{A\subset A_j\setminus A_{j-1}\\ A \text{ non-critical}}} 2^{2G_jN}2^{-\gamma(A)N}p_N^{(2)}(j,\mathfrak{R},A,\varepsilon_2)\\
&\overset{\text{Lemma 1.22}}{\lesssim} \sum_{\substack{A\subset A_j\setminus A_{j-1}\\ A \text{ critical}}} e^{-const\times\varepsilon_1\sqrt{N}}+\sum_{\substack{A\subset A_j\setminus A_{j-1}\\ A \text{ non-critical}}} 2^{-\gamma(A)N}\exp\Big\{\beta_j^2\widehat{\alpha}_j(A)\Big[1-\frac{1}{2}(1-\varepsilon_2)^2\Big]N\Big\}.
\end{aligned}
\tag{1.48}
$$

For non-critical A, $\beta_j^2\widehat{\alpha}_j(A)<\gamma(A)2\log 2$ strictly, so we can find ε_2 small enough such that

$$
\delta'(\varepsilon_1)\overset{\text{def}}{=}\max_{j\le K}\max_{A\subsetneq A_j\setminus A_{j-1};\, A \text{ non-critical}}\left\{\beta_j^2\widehat{\alpha}_j(A)\Big[1-\frac{1}{2}(1-\varepsilon_2)^2\Big]-\gamma(A)\log 2\right\}<0.
\tag{1.49}
$$

The second sum on the r.h.s of (1.48) is thus $\lesssim \exp\big[-|\delta'|N\big]$, while the first sum is $\lesssim \exp\big[-const\times\varepsilon_1\sqrt{N}\big]$. This proves the claim. $\square$

Suppose now that two configurations $\sigma,\tau\in\Sigma_{N,A_j}^{R,\varepsilon_1,\varepsilon_2}$ are such that $\sigma_s=\tau_s$ for some $s\in A_m\setminus A_{m-1}$ for some $m\le j$ but $\sigma_t\neq\tau_t$ for some $t\in A_r\setminus A_{r-1}$ and $r<m$. We next address the case when there are numbers k,l,m, $0\le k<l<m\le j$ such that $\sigma_{A_k}=\tau_{A_k}$, $\sigma_r\neq\tau_r$ $\forall r\in A_l\setminus A_k$, and $\sigma_{A_m\setminus A_l}=\tau_{A_m\setminus A_l}$.

Proposition 1.24 (Propagation)**.** *For positive ε_1 and small enough ε_2 there exists positive const such that*

$$
\mathbb{P}\left[\exists\,\sigma,\tau\in\Sigma_{N,A_m}^{R,\varepsilon_1,\varepsilon_2}:\sigma_{A_k}=\tau_{A_k},\sigma_r\neq\tau_r\,\forall r\in A_k\setminus A_l,\ \sigma_{A_m\setminus A_l}=\tau_{A_m\setminus A_l}\right]\lesssim e^{-const\times N}.
\tag{1.50}
$$

Proof. Without loss of generality we may assume $m=l+1$. Consider two configurations $\sigma,\tau\in\Sigma_{N,A_{l+1}}$ which differ on the whole $A_l\setminus A_k$ but $\sigma_{A_{l+1}\setminus A_l}=\tau_{A_{l+1}\setminus A_l}$. By the irreducibility condition $\boldsymbol{\mathcal{I}}_2$ there exists $J\in\mathcal{P}_{A_{l+1}}\setminus\mathcal{P}_{A_l}$ such that $\sigma_J\neq\tau_J$ in which case there must be a strict subset $A\subsetneq A_{l+1}\setminus A_l$ such that $\sigma_J=\tau_J$ for all $J\in\widehat{\mathcal{P}}_{l+1,A}$ and $\sigma_J\neq\tau_J$ for all $J\in\widehat{\mathcal{P}}^c_{l+1,A}$ (loosely speaking, the associated random variables $\overline{X}_{\sigma(1),\ldots,\sigma(l+1)}$

and $\overline{X}_{\tau(1),\dots,\tau(l+1)}$ cannot coincide). We can therefore bound the l.h.s. of (1.50) by

$$\sum_{\substack{A\subsetneq A_{l+1}\setminus A_l\\ A\text{ critical}}} \sum^{*} \mathbb{P}\Big[\overline{X}_{\sigma(1),\dots,\sigma(j)} \text{ and } \overline{X}_{\tau(1),\dots,\tau(j)} \in \mathfrak{R} \quad \text{for all } j=1,\dots k,\dots,l+1;$$

$$\mathbf{T}_1(\sigma,l,A,\varepsilon_1) \text{ and } \mathbf{T}_1(\tau,l,A,\varepsilon_1) \text{ hold}\Big]+$$

$$\sum_{\substack{A\subsetneq A_{l+1}\setminus A_l\\ A\text{ non-critical}}} \sum^{*} \mathbb{P}\Big[\overline{X}_{\sigma(1),\dots,\sigma(j)} \text{ and } \overline{X}_{\tau(1),\dots,\tau(j)} \in \mathfrak{R} \quad \text{for all } j=1,\dots l+1;$$

$$\mathbf{T}_2(\sigma,l,A,\varepsilon_2) \text{ and } \mathbf{T}_2(\tau,l,A,\varepsilon_2) \text{ hold}\Big], \tag{1.51}$$

where $\sum^{\star}$ runs over those σ,τ in $\Sigma_{N,A_{l+1}}$ such that $\sigma_J=\tau_J$ for all $J\in\widehat{\mathcal{P}}_{l+1,A}$, $\sigma_J\neq\tau_J$ $J\in\widehat{\mathcal{P}}^c_{l+1,A}$, $\sigma_{A_k}=\tau_{A_k}$, $\sigma_s\neq\tau_s\,\forall s\in A_l\setminus A_k$, $\sigma_{A_{l+1}\setminus A_l}=\tau_{A_{l+1}\setminus A_l}$.

We also observe that $\sigma_s\neq\tau_s$ for all $s\in A_l\setminus A_k$ implies that the random variables $\overline{X}_{\sigma(1),\dots,\sigma(j)}$ and $\overline{X}_{\tau(1),\dots,\tau(j)}$ are independent for all $j=k+1\dots l$. In fact, for every $J\in\mathcal{P}_{A_l}\setminus\mathcal{P}_{A_k}$ by construction $J\cap(A_l\setminus A_k)\neq\emptyset$; this amounts to say that for every such J there exists at least one $s\in A_l\setminus A_k$ with $J\ni s$.

The above remarks, together with some simple counting steadily yield

$$\begin{aligned}
(1.51) &\lesssim 2^{N\left[\gamma(A_k)+2\gamma(A_l\setminus A_k)+\gamma(A_{l+1}\setminus A_l)\right]}\prod_{r\le k} p_N\big(r,\mathfrak{R}\big)\prod_{r=k+1}^{l} p_N\big(r,\mathfrak{R}\big)^2\times\\
&\qquad\times\Bigg\{\sum_{\substack{A\subsetneq A_{l+1}\setminus A_l\\ A\text{ critical}}} p_N^{(2,crit)}(l+1,\mathfrak{R},A,\varepsilon_1)+\sum_{\substack{A\subsetneq A_{l+1}\setminus A_l\\ A\text{ non-critical}}} p_N^{(2)}\big(l+1,\mathfrak{R},A,\varepsilon_2\big)\Bigg\}\\
&\overset{\text{Lemma 1.22}}{\lesssim} \sum_{\substack{A\subsetneq A_{l+1}\setminus A_l\\ A\text{ non-critical}}} \exp\left\{2\log 2G_{l+1}N\left[\Big(1-\frac{1}{2}\big(1-\varepsilon_2\big)^2\Big)\frac{\widehat{\alpha}_{l+1}(A)}{\Delta_{l+1}}-\frac{1}{2}\right]\right\}+\\
&\qquad\qquad+\sum_{\substack{A\subsetneq A_{l+1}\setminus A_l\\ A\text{ critical}}} 2^{(\gamma(A)-G_{l+1})N}\exp\big[-\mathit{const}\times\varepsilon_1\sqrt{N}\big].
\end{aligned}\tag{1.52}$$

Clearly, the second sum on the r.h.s above is $\lesssim \exp\big[-\big|\delta'\big|N\big]$ for

$$\delta' \overset{\text{def}}{=} \max_{l\le K-1}\max_{\substack{A\subsetneq A_{l+1}\setminus A_l\\ A\text{ critical}}}\Big\{\gamma(A)-G_{l+1}\Big\}<0.$$

It is crucial that the first sum runs over (non-critical) subsets strictly included in $A_l \backslash A_{l+1}$, since it guarantees that $\max_{A \subsetneq A_{l+1} \backslash A_l} \widehat{\alpha}_{l+1}(A) < \Delta_{l+1}$ and thus, for small enough ε_2,

$$\delta''(\varepsilon_2) \stackrel{\text{def}}{=} \max_{l \leq K-1} \max_{A \subsetneq A_{l+1} \backslash A_l} \left\{ (2 \log 2) G_{l+1} \left[\left(1 - \frac{1}{2}(1-\varepsilon_2)^2\right) \frac{\widehat{\alpha}_{l+1}(A)}{\Delta_{l+1}} - \frac{1}{2} \right] \right\} < 0. \quad (1.53)$$

This settles the Lemma with *const* $\stackrel{\text{def}}{=} \min\{|\delta'|, |\delta''|\}$. □

Proof of Proposition 1.8. We fix $\epsilon > 0$. It is easily seen that there exists $\mathsf{N} = \mathsf{N}(\epsilon)$ such that the probability that there exist more than N configurations in $\Sigma_{N,A_j}^{R,\varepsilon_1,\varepsilon_2}$ is smaller than $\epsilon/2$ (this follows from Markov inequality together with the estimates from Lemma 1.21). Therefore, it suffices to estimate the probability that, out of a finite number N of configurations in $\Sigma_{N,A_j}^{R,\varepsilon,\varepsilon_2}$ some of them form a non ultrametric couple. But this case is taken care of by Proposition 1.23 and 1.24 (and a straightforward combination of the two). By choosing ε_2 *small enough*, in the range of validity of (1.49) and (1.53), the probability of such an event is of order $\exp[-const \times \sqrt{N}]$, thus smaller than $\epsilon/2$ for large enough N. □

1.3.4. Weak limits of the energy levels. The main goal of this section is to prove the multivariate Poisson Approximation as explained in the considerations following the statement of Proposition 1.8, and then to derive from this the weak convergence of $X_{N,j}^{R,\varepsilon_1,\varepsilon_2}$.

We begin with a technical estimate. For bounded real subset $\mathfrak{M}$, and $\delta, \rho > 0$ we set:

$$p_N^{\delta,\rho}(j, \mathfrak{M}) \stackrel{\text{def}}{=} \mathbb{P}\Bigg[\overline{Y}_j \in \mathfrak{M};\ \forall \text{ critical } B \subsetneq A_j \setminus A_{j-1} : \frac{Y_{j,B}}{\widehat{\alpha}_j(B)} - \frac{Y_{j,B}^c}{\widehat{\alpha}_j^c(B)} \leq -\delta;$$
$$\forall\, A \subset A_j \setminus A_{j-1}, \widehat{\alpha}_j(A) > 0 : \overline{Y}_{N,j}(A) \leq \beta_j (1+\rho) \widehat{\alpha}_j(A) \sqrt{N}\Bigg].$$

Lemma 1.25. *For $N \uparrow \infty$, it holds*

$$p_N^{\delta,\rho}(j, \mathfrak{M}) = \mathcal{C}_{j,\delta} \times 2^{-G_j N} \int_{\mathfrak{M}} \beta_j \exp\left[-\beta_j x + o(1)\right] dx + O\big(2^{-G_j N} e^{-const \times N}\big)$$

Proof. Clearly,

$$\begin{aligned} p_N^{\delta,\rho}(j, \mathfrak{M}) &= \mathbb{P}\Big[\overline{Y}_j \in \mathfrak{M};\ \forall \text{ critical } B \subsetneq A_j \setminus A_{j-1} : \frac{Y_{j,B}}{\widehat{\alpha}_j(B)} - \frac{Y_{j,B}^c}{\widehat{\alpha}_j^c(B)} \leq -\delta\Big] + \\ &\quad - \mathbb{P}\Big[\overline{Y}_j \in \mathfrak{M};\ \exists\, A \subset A_j \setminus A_{j-1}, \widehat{\alpha}_j(A) > 0 : \overline{Y}_{N,j}(A) > \beta_j (1+\rho) \widehat{\alpha}_j(A) \sqrt{N}\Big] \\ &= (I) - (II). \end{aligned} \quad (1.54)$$

As for (I), we claim that, somewhat surprisingly, the random variable $\overline{Y}_j = \sqrt{N} Y_j - a_{N,j}$ is independent of the collection $\Big(\frac{Y_{j,B}}{\widehat{\alpha}_j(B)} - \frac{Y_{j,B}^c}{\widehat{\alpha}_j^c(B)}; B \subsetneq A_j \setminus A_{j-1} \text{ is critical}\Big)$. This is best seen by inspection of the covariance: for critical B, since $Y_j = Y_{j,B} + Y_{j,B}^c$, we have

$$\mathbb{E}\left[Y_j \cdot \left(\frac{Y_{j,B}}{\widehat{\alpha}_j(B)} - \frac{Y_{j,B}^c}{\widehat{\alpha}_j^c(B)}\right)\right] = \frac{1}{\widehat{\alpha}_j(B)}\mathbb{E}\big[Y_{j,B}^2\big] - \frac{1}{\widehat{\alpha}_j^c(B)}\mathbb{E}\big[(Y_{j,B}^c)^2\big] = 0,$$

and thus $(I) = \mathcal{C}_{j,\delta} \times p_N(j,\mathfrak{M})$ exactly. On the other hand,

$$0 \leq (II) \leq \sum_{A \subset A_j \setminus A_{j-1}, \widehat{\alpha}_j(A)>0} p_N^{>}(j,\mathfrak{M},A,\rho).$$

The Lemma then obviously follows by the asymptotics established in Lemma 1.21. $\square$

We may now move to the multivariate Poisson approximation. We stress that henceforth, we think of $R, \varepsilon_1, \varepsilon_2 > 0$ as being given, with ε_2 *small enough*, in the range of validity of (1.49) and (1.53).

Let us take $\sigma^1, \ldots, \sigma^r \in \Sigma_{N,A_{j-1}}$ distinct reference configurations, with overlaps in the chain $\mathbf{T}^{j-1}$ (this already entails that $\sigma_i^p \neq \sigma_i^q$ for all $i \in A_{j-1} \setminus A_{j-2}$, $p, q = 1, \ldots, r, p \neq q$). Consider also $B_1, \ldots, B_r \subset [-R, R]$, and the random vector:

$$\left(\sum^{(1)} \delta_{\overline{X}_{\sigma^1,\sigma(j)}}(B_1), \ldots, \sum^{(k)} \delta_{\overline{X}_{\sigma^k,\sigma(j)}}(B_r)\right),$$

where the rth-sum $\sum^{(r)}$ runs over those $\sigma(j) \in \Sigma_{N,A_j \setminus A_{j-1}}$ such that $\mathbf{T_1}(\varepsilon_1)$ and $\mathbf{T_2}(\varepsilon_2)$ hold for all critical (resp. non-critical) subsets of $A_j \setminus A_{j-1}$. Remark that by Lemma 1.25,

$$\lim_{N\to\infty} \mathbb{E}\left[\sum^{(r)} \delta_{\overline{X}_{\sigma^r,\sigma(j)}}(B_r)\right] = \lim_{N\to\infty} 2^{G_j N} p_N^{\varepsilon_1,\varepsilon_2}(j, B_r) = \int_{B_r} \mathcal{C}_{j,\varepsilon_1} \beta_j \exp\Big[-\beta_j t\Big] dt = \mu_{\varepsilon_1}(B_r).$$

According to [**2**, p. 236], the multivariate Poisson convergence is equivalent to weak convergence of the sum of the vector's componenent, $V_N \stackrel{\text{def}}{=} \sum_{r=1}^k \sum^{(r)} \delta_{\overline{X}_{\sigma(1),\ldots,\sigma(j)}}(B_r)$, towards a Poisson random variable, say V, of parameter $\sum_{r=1}^k \mu(B_r)$. The latter, however, follows steadily through the *Chen-Stein's method*, cfr. [**2**]. We sketch the main steps.

Introduce the index set

$$\Gamma \stackrel{\text{def}}{=} \left\{(r, \sigma^r, \sigma(j)) : r = 1, \ldots, k,\ \sigma(j) \in \Sigma_{N,A_j \setminus A_{j-1}}^{R,\varepsilon_1,\varepsilon_2}\right\}.$$

For given $\alpha = (r, \sigma^r, \sigma) \in \Gamma$, consider the subset $\Gamma_\alpha \subset \Gamma$ consisting of those $(q, \sigma^q, \tau) \in \Gamma$ with the random variables $\overline{X}_{\sigma^r,\sigma}$ and $\overline{X}_{\sigma^q,\tau}$ such that $\mathbb{E}\left(\overline{X}_{\sigma^r,\sigma}\overline{X}_{\sigma^q,\tau}\right) \neq a_{N,j}^2$, that is they are correlated. (In the classical Chen-Stein terminology, Γ_α is the "weak dependency neighborhood" of the index α.) We set

$$p_\alpha \stackrel{\text{def}}{=} \mathbb{P}\Big[\overline{X}_{\sigma^r,\sigma} \in B_r, (\sigma^r, \sigma) \text{ satisfies truncation } \mathbf{T}_1(\varepsilon_1), \mathbf{T}_2(\varepsilon_2)\Big]$$

and define $Z_\alpha \stackrel{\text{def}}{=} \sum^\star_{(q,\sigma^q,\tau)\in\Gamma_\alpha} \delta_{\overline{X}_{\sigma^q,\tau}}(B_q)$, the sum running over those configurations satisfying condition $\mathbf{T_1}(\varepsilon_1)$ and $\mathbf{T_2}(\varepsilon_2)$. According to the *Chen-Stein bound*, cfr. [**2**, Theorem 1.A], the total variation distance between V_N and V is bounded above by

$$\sum_\alpha \Big\{ p_\alpha^2 + \sum_{\alpha'\in\Gamma_\alpha} p_\alpha p_{\alpha'} \Big\} + \sum_{\alpha=(r,\sigma^r,\tau)\in\Gamma} \mathbb{E}[\delta_{\overline{X}_{\sigma^r,\tau}}(B_r) 1_{\mathbf{T_1},\mathbf{T_2}\text{ are satisfied}} \times Z_\alpha]. \tag{1.55}$$

Writing things out, one immediate realizes that exactly the same terms as in Proposition 1.23 make their appearance in expression (1.55). (These terms are in fact taken care of by Lemma 1.21.) Here is the outcome of the considerations: the first sum is of order $\exp(-const \times N)$ for some positive *const*, while the second sum is bounded, *mutatis mutandis*, by a constant times the l.h.s of (1.47). The total variation distance between V_N and V is therefore of order $\exp(-const \times \varepsilon_1\sqrt{N})$. Letting $N \to \infty$ yields the Poisson convergence.

We now move to the proof of the weak convergence of the process associated to the energy levels, $\mathcal{N}^{R,\varepsilon_1,\varepsilon_2}_{j,N}$. We start with the case $j=1$ and sketch the main principles: for $B \subset [-R,R]$, by the Poisson convergence we have

$$\lim_{N\to\infty} \mathbb{P}\Big[\sharp\{\sigma \in \Sigma^{R,\varepsilon_1,\varepsilon_2}_{N,A_1} : \overline{X}_{\sigma_{A_1}} \in B\} = L\Big] = e^{-\mu(B)}\frac{\mu(B)^L}{L}. \tag{1.56}$$

On the other hand

$$\Big\{\sharp\{\sigma \in \Sigma^{R,\varepsilon_1,\varepsilon_2}_{N,A_1} : \overline{X}_{\sigma_{A_1}} \in B\} = L\Big\} =$$
$$\Big\{\mathcal{N}^{R,\varepsilon_1,\varepsilon_2}_{N,j}(B\times B;\emptyset) = \frac{L(L-1)}{2}\Big\} \bigcup \left\{ \begin{array}{l} \text{there exists } L \text{ configurations falling into} \\ B, \text{ some form non-ultrametric couples.} \end{array} \right\}. \tag{1.57}$$

In virtue of Proposition 1.8, the $\mathbb{P}$-probability of the second event is in the limit $N\to\infty$ irrelevant, thus

$$\lim_{N\to\infty} \mathbb{P}\Big[\mathcal{N}^{R,\varepsilon_1,\varepsilon_2}_{N,j}(B\times B;\emptyset) = L(L-1)/2\}\Big] = e^{-\mu(B)}\frac{\mu(B)^L}{L!}.$$

This line of reasoning carries over to more involved events (remark that there are no integrability issues to be addressed: on the one hand, the involved subsets must all be bounded, and second, tightness is easily seen to follow from Markov inequality), and obviously for the limiting picture as well.
As for the induction step $(j-1)\to j$, we lighten notation omitting the superscripts $R,\varepsilon_1,\varepsilon_2$. It pays to introduce the projection $\mathfrak{P}:\mathbb{R}^2\to\mathbb{R}, (x,y)\mapsto x+y$, and to consider the points

$$\Big\{\Big(\widehat{X}_{\sigma(1),\dots,\sigma(j-1)}, \overline{X}_{\sigma(1),\dots,\sigma(j-1),\sigma(j)}\Big), \sigma \in \Sigma^{R,\varepsilon_1,\varepsilon_2}_{N,A_j}\Big\}$$

This induces naturally a process $\mathcal{N}^{(2)}_{N,j} \in \mathcal{M}_{mp}\Big((\mathbb{R}^2)^{(2)} \times 2^{A_j}\Big)$. The process $\mathcal{N}_{N,j}$ is then the "image" of $\mathcal{N}^{(2)}_{N,j}$ under the projection $\mathfrak{P}$ (the points $(\widehat{X}_{\sigma(1),\dots,\sigma(j-1)}, \overline{X}_{\sigma(1),\dots,\sigma(j)})$ are projected to $\widehat{X}_{\sigma(1),\dots,\sigma(j-1)}+\overline{X}_{\sigma(1),\dots,\sigma(j)}=\widehat{X}_{\sigma(1),\dots,\sigma(j)})$. As for the finite dimensional distributions of the "multidimensional process" $\mathcal{N}^{(2)}_{N,j}$, again by Proposition 1.8, for bounded $B,\mathfrak{M}\subset\mathbb{R}$ we have that

$$\lim_{N\to\infty}\mathbb{P}\Big[\mathcal{N}^{(2)}_{N,j}(\mathbb{R}\times\mathbb{R};A)>0\Big]=0,\ \forall A\in 2^{A_j}\setminus\{\emptyset,A_1,\dots,A_{j-1}\}.$$

As for the events involving overlaps in the chain $\{\emptyset,\dots,A_j\}$, they are easily handled through the following remark: conditionining the process $\mathcal{N}^{(2)}_{N,j}$ to the sigma-field generated by the process $\mathcal{N}_{N,j-1}$ amounts to prescribe a finite number, say L, of configurations $\sigma^1,\dots,\sigma^L\in\Sigma_{N,A_{j-1}}$, as well as their overlap structure. By ultrametricity, the overlaps among these L configurations take values in the chain $\{\emptyset,\dots,A_{j-1}\}$ only. But then, it is easy to reformulate the finite dimensional distributions of the process $\mathcal{N}^{(2)}_{N,j}$ given the process $\mathcal{N}_{N,j-1}$ into finite dimensional probabilities of the point processes $(\overline{X}_{\sigma^r,\tau},\tau\in\Sigma_{N,A_j\setminus A_{j-1}})$, with prescribed $\sigma^1,\dots,\sigma^L$ for $r=1,\dots,L$, thus reducing the problem to the $j=0$ case described above. On this level, the mechanism of *suppression of structures* provides the link with the convergence towards multivariate Poisson random variables (there is in fact no possible "partial overlap", only $A_j\setminus A_{j-1}$ comes in question - as explained in (1.57)). So, $\mathcal{N}^{(2)}_{N,j}$ converges weakly to the process $\mathcal{N}^{(2)}_j$ naturally induced by the points $\big\{(y_{\mathbf{i}},y_{\mathbf{i},l});\mathbf{i}\in\mathbb{N}^{j-1},l\in\mathbb{N}\big\}$ on $\mathbb{R}^2$. By continuity of the projection $\mathfrak{P}:[-(j-1)R,(j-1)R]\times[-R,R]\to\mathbb{R},(x_1,x_2)\mapsto x_1+x_2$, weak convergence of $\mathcal{N}_{N,j}$ then follows.

1.3.5. Localization of the Gibbs measure.

Proof of Lemma 1.13. This is essentially an adaptation of [**6**, Lemma 3.1] to our more general setting, so we only sketch the main differences. We begin with the easy observation

$$\mathbb{E}\big[Z_\sigma\big]=\exp\left[\sum_{j=m+1}^{K}\frac{\beta^2}{2}\Delta_jN+NG_j\log 2\right],$$

and also introduce some notation. For $A\subset(I\setminus A_m),\tau\in\Sigma_N$ and $\varepsilon>0$ we set

$$X_\tau(A)\stackrel{\text{def}}{=}\sum_{J\in\widehat{\mathcal{P}}_{A,m}}X^J_{\tau_J},\quad \widehat{Z}_\sigma\stackrel{\text{def}}{=}\widehat{\sum}\exp\Big[\beta\big(X_{\tau(1),\dots,\tau(m+1)}+\dots+X_{\tau(1),\dots,\tau(K)}\big)\Big],$$

where $\widehat{\sum}$ runs over those $\tau\in\Sigma_N$ such that $\tau_{A_m}=\sigma$ and for all $A\subseteq I\setminus A_m$ the random variables $X_\tau(A)$ are bounded by $(\beta+\varepsilon)\widehat{\alpha}_m(A)N$. We proceed to show that the claim of the Lemma holds, at least for small enough ε. We first write

$$\frac{Z_\sigma}{\mathbb{E}[Z_\sigma]}=\frac{\widehat{Z}_\sigma}{\mathbb{E}[\widehat{Z}_\sigma]}\times\frac{\mathbb{E}[\widehat{Z}_\sigma]}{\mathbb{E}[Z_\sigma]}+\frac{Z_\sigma-\widehat{Z}_\sigma}{\mathbb{E}[Z_\sigma]}=(I)\times(II)+(III).$$

It is easily seen that to $\varepsilon>0$ one can find $\eta>0$ such that $1-e^{-\eta N}\leq(II)\leq 1$, for N large enough. This, together with Markov inequality entails that $\mathbb{P}\left[(III)\geq e^{-\eta N/2}\right]\lesssim e^{-\eta N/2}$.

Therefore, on a set of $\mathbb{P}$-probability exponentially close to unity, the following holds:

$$\frac{Z_\sigma}{\mathbb{E}[Z_\sigma]} = (I) \times \left\{1 - O(e^{-constN})\right\} + O\left(e^{-constN}\right), \tag{1.58}$$

for $N \to \infty$ and some *const* > 0 whose precise value is not important. In particular, we see from (1.58) that the claim of the Lemma follows as soon as we prove that for some $\delta_1, \delta_2 \in (0,1)$

$$\mathbb{P}\left[\left|\log(I)\right| \geq N^{-\delta_1}\right] \lesssim \exp\left[-N^{\delta_2}\right]. \tag{1.59}$$

To see the latter, let us fix $\delta_1 \in (0,1)$. We write:

$$\begin{aligned}
&\mathbb{P}\left[\left|\log(I)\right| \geq N^{-\delta_1}\right] \\
&\quad = \mathbb{P}\Big[(I) \geq \exp(N^{-\delta_1}) \text{ or } (I) \leq \exp(-N^{-\delta_1})\Big] \\
&\quad = \mathbb{P}\left[\Big((I)-1\Big)^2 \geq (\exp(N^{-\delta_1})-1)^2 \text{ or } \Big((I)-1\Big)^2 \geq (\exp(-N^{-\delta_1})-1)^2\right] \\
&\quad \leq \mathbb{P}\left[\Big((I)-1\Big)^2 \geq \min\Big\{(\exp(N^{-\delta_1})-1)^2; (\exp(-N^{-\delta_1})-1)^2\Big\}\right] \\
&\quad \overset{\text{(Markov)}}{\leq} \frac{1}{m(N,\delta_1)} \frac{\mathbb{E}\Big[(\widehat{Z}_\sigma - \mathbb{E}[\widehat{Z}_\sigma])^2\Big]}{\mathbb{E}\big[\widehat{Z}_\sigma\big]^2},
\end{aligned} \tag{1.60}$$

with $m(N,\delta_1) \overset{\text{def}}{=} \min\left\{(\exp(N^{-\delta_1})-1)^2; (\exp(-N^{-\delta_1})-1)^2\right\}$. It is now crucial that $\beta < \beta_{m+1}$ strictly: this ensures that for ε small enough (recall the construction of the chain $\mathbf{T}$) we have

$$\eta' \overset{\text{def}}{=} \inf_{A \subset (I \setminus A_m)} \left\{\gamma(A)\log 2 - \left[\beta^2 - \frac{(\beta-\varepsilon)^2}{2}\right]\widehat{\alpha}_m(A)\right\} > 0. \tag{1.61}$$

Given this, expanding the square in the numerator of the r.h.s of (1.60) and exploiting the usual bounds on gaussian integrals yields

$$\begin{aligned}
\mathbb{P}\Big[\left|\log(I)\right| \geq N^{-\delta_1}\Big] &\lesssim \frac{1}{m(N,\delta_1)} \sum_{A \subset (I\setminus A_m)} 2^{-\gamma(A)N} \exp\left[N\big(\beta^2 - \frac{(\beta-\varepsilon)^2}{2}\big)\widehat{\alpha}_m(A)\right] \\
&\overset{(1.61)}{\lesssim} \frac{\exp\big[-\eta' N\big]}{m(N,\delta_1)},
\end{aligned} \tag{1.62}$$

which is clearly more than needed to get (1.59). Lemma 1.13 then easily follows. □

To prove Proposition 1.12 some additional a priori estimates are required. For notational simplicity we set $a_N \overset{\text{def}}{=} \sum_{j\leq m} a_{N,j} + \sum_{j=m+1}^K \frac{\beta}{2}\Delta_j N + G_j N \log 2/\beta$.

Lemma 1.26. *Let $\epsilon > 0$. There exists positive ϕ such that for every $j \leq m$*

$$\mathbb{P}\left[\sum_{\exists j \leq m: \widehat{X}_{\sigma(1),\dots,\sigma(j)} \leq -\phi} \exp\Big[\beta\big(X_\sigma - a_N\big)\Big] \geq \epsilon\right] \leq \epsilon. \tag{1.63}$$

PROOF. By Proposition 1.5 we can find a constant C such that for large enough N

$$\mathbb{P}\Big[\forall j \le m, \forall \tau \in \Sigma_{N,A_j}\ \widehat{X}_{\tau(1),\dots,\tau(j)} \le C\Big] \ge 1-\epsilon/2,$$

in which case the l.h.s of (1.63) is then bounded by $\mathbb{P}\Big[\widehat{\sum}\exp\big[\beta(X_\sigma - a_N)\big] \ge \epsilon\Big] + \epsilon/2$, with $\widehat{\sum}$ running over those $\sigma \in \Sigma_N$ such that $\widehat{X}_{\sigma(1),\dots,\sigma(l)} \le C$ for all $l = 1,\dots,m$ but $\widehat{X}_{\sigma(1),\dots,\sigma(j)} \le -\phi$ for some $j = 1,\dots,m$. We have:

$$\begin{aligned}
&\mathbb{P}\Big[\widehat{\sum}\exp\big[\beta(X_\sigma - a_N)\big] \ge \epsilon\Big] \le \\
&\le \epsilon^{-1} \sum_{\substack{\sigma\in\Sigma_N \\ j=1,\dots,m}} \mathbb{E}\Big[\exp\big[\beta(X_\sigma - a_N)\big]; \forall l \le m : \widehat{X}_{\sigma(1),\dots,\sigma(l)} \le C,\ \widehat{X}_{\sigma(1),\dots,\sigma(j)} \le -\phi\Big] \\
&\le \epsilon^{-1} 2^{\gamma(A_m)N} \sum_{j=1}^{m} \mathbb{E}\Big[\exp\big[\beta \widehat{Y}_m\big];\ \forall l \le m : \widehat{Y}_l \le C,\ \text{but } \widehat{Y}_j \le -\phi\Big] \\
&\lesssim \epsilon^{-1} \sum_{j\le m} \exp\Big[\sum_{l\ne j}(\beta_{l+1}-\beta_l)C - \big(\beta_{j+1}-\beta_j\big)\phi + o(1)\Big]
\end{aligned} \tag{1.64}$$

(the first step above by Markov inequality, the second by simply integrating out the unrestricted random variables $X_{\sigma(1),\dots,\sigma(l)}$ ($l = m+1,\dots,K$) and the third by Lemma 1.19). It thus suffices to choose ϕ large enough in the positive to have that (1.64) is smaller than $\epsilon/2$, settling the proof of the Lemma. □

PROOF OF PROPOSITION 1.12. We claim that to arbitrary $\epsilon > 0$ there exists $\widehat{C} > 0$ such that

$$\mathbb{P}\left[\mathcal{G}_{\beta,N}\left(\exists j \le m :\ \widehat{X}_{\sigma(1),\dots,\sigma(j)} \notin \left(-\widehat{C},\widehat{C}\right)\right) \ge \epsilon\right] \le \epsilon. \tag{1.65}$$

This obviously implies that there exist $\overline{C} > 0$ such that the claim of Proposition 1.12 holds. To see (1.65), we first modify the definition of the Gibbs measure slightly, subtracting the constant βa_N to the energies: $\mathcal{G}_{\beta,N}(\sigma) = \exp\big[\beta(X_\sigma - a_N)\big]/Z_{a_N}(\beta)$ with $Z_{a_N}(\beta) \stackrel{\text{def}}{=} \sum_{\tau\in\Sigma_N}\exp\big[\beta(X_\tau - a_N)\big]$. We now claim that to given ϵ there exists $\eta > 0$ such that, for N large enough

$$\mathbb{P}\Big[Z_{a_N}(\beta) \le \eta\Big] \le \frac{\epsilon}{2}. \tag{1.66}$$

The l.h.s above is to any $R > 0$ evidently bounded by

$$\mathbb{P}\left[\widehat{\sum}_R \exp\Big[\beta\big(\widehat{X}_{\sigma(1),\dots,\sigma(m)} + \frac{1}{\beta}\log\frac{Z_{\sigma(1),\dots,\sigma(m)}}{\mathbb{E}[Z_{\sigma(1),\dots,\sigma(m)}]}\big)\Big] \le \eta\right]$$

with $\widehat{\sum}_R$ running over those $\sigma \in \Sigma_{N,A_m}$ only such that $\widehat{X}_{\sigma(1),\dots,\sigma(m)} \in (-R,R)$. It is also easily seen that to any $\epsilon' > 0$ this sum runs over at most $\mathsf{N} = \mathsf{N}(\epsilon')$ configurations with $\mathbb{P}$-probability greater than $(1-\epsilon')$. By Lemma 1.13 the contributions of each single term $\log\big(Z_\sigma/\mathbb{E}[Z_\sigma]\big)$ associated to these N configurations is the limit $N \to \infty$ vanishing (with overwhelming probability). It is therefore sufficient to prove that to every $\tilde{\epsilon}$ there exist $\tilde{\eta}$ such that

$$\mathbb{P}\left[\widehat{\sum}_R \exp\left[\beta \widehat{X}_{\sigma(1),\dots,\sigma(m)}\right] \leq \tilde{\eta}\right] \leq \frac{\tilde{\epsilon}}{2}.$$

This is however straightforward, since for $x < -R$

$$\widehat{\sum}_R \exp\left[\beta \widehat{X}_{\sigma(1),\dots,\sigma(m)}\right] \leq \exp(\beta x) \Longrightarrow \sharp\Big\{\sigma \in \Sigma_{N,A_m} : \ \widehat{X}_{\sigma(1),\dots,\sigma(m)} \geq -R\Big\} = 0. \tag{1.67}$$

In virtue of Theorem 1.9.b) and the properties of the limiting point process $\widehat{X}_m$, it is plain that the probability of the event on the r.h.s above can be made (for large enough N) as small as needed by simply choosing R large enough in the positive. On the other hand, by Proposition 1.5 and Lemma 1.26, to given $\eta, \epsilon > 0$ we can find positive $\widehat{C}$ such that

$$\mathbb{P}\left[\sum_{\sigma\in\Sigma_N;\, \exists j\leq m:\, \widehat{X}_{\sigma(1),\dots,\sigma(j)}\notin(-\widehat{C},\widehat{C})} \exp\left[\beta(X_\sigma - a_N)\right] \geq \eta\epsilon\right] \leq \frac{\epsilon}{2},$$

which together with (1.66) yields (1.65) and thus settles the proof of Proposition 1.12.

□

Proof of Theorem 1.4. The main issue is to prove that the normalization $\mathcal{N}$ commutes with taking the $N \to \infty$ limit. The rest is implied by the remarkable properties of the PPP considered.

The low temperature. Recall that $\Xi_{\beta,N}$ is the law on $\mathcal{M}_{mp}\left((\mathbb{R}^+)^{(2)} \times 2^I\right)$ naturally induced by the collection of points $(\exp[\beta(X_\sigma - a_N)]\big/Z_{a_N}(\beta), \sigma \in \Sigma_N)$. Set then $H_{N,K} \stackrel{\text{def}}{=} (\exp\left[\beta(X_\sigma - a_N)\right], \sigma \in \Sigma_N)$. This is nothing else than the image of the PP of the energy levels under the mapping $\exp(\beta\cdot)$, in which case *(cfr.* [**4**, Prop. 8.5] *and a straightforward generalization)* it follows by Theorem 1.9 that $H_{N,K}$ converges weakly to a PP $H_K \stackrel{\text{def}}{=} (\eta_{\mathbf{i}}, \mathbf{i} \in \mathbb{N}^K)$ with $\eta_{\mathbf{i}} = \eta^1_{\mathbf{i}_1}\eta^2_{\mathbf{i}_2}\cdots\eta^K_{\mathbf{i}_K}$ and the following properties: For $l \leq K$ and multi-index $\mathbf{i}_{l-1}$, the point process $(\eta^l_{\mathbf{i}_{l-1},i_l};\ i_l \in \mathbb{N})$ is poissonian with density $\mathcal{C}_l x_l(\beta) \cdot t^{-x_l(\beta)-1}dt$ on $\mathbb{R}^+$; The η^l are independent for different l; $(\eta^l_{\mathbf{i}_{l-1},i_l};\ i_l \in \mathbb{N})$ are independent for different $\mathbf{i}_{l-1}$. Given such a PP, it is easily seen that $\sum_{\mathbf{i}} \eta_{\mathbf{i}} < \infty$ almost surely. (This is mainly due to the fact that $x_1(\beta) < x_2(\beta) < \cdots < x_K(\beta)$. For more on this, cfr. [**4**, Prop. 9.5] and a straightforward generalization.) We may thus consider the new collection of *normalized* points given by $(\overline{\eta_{\mathbf{i}}}; \mathbf{i} \in \mathbb{N}^K)$, which induces naturally an element of $\mathcal{M}_{mp}\left((\mathbb{R}^+)^{(2)} \times 2^I\right)$ with possible marks those from the chain $\mathbf{T} = \{A_0, A_1, \dots, A_K\}$ only. We denote by Ξ_β its law.

With the new notation $Z_{a_N}(\beta) = \int xH_{N,K}(dx)$, and by Proposition 1.5 and Lemma 1.26 we have that to $\epsilon > 0$ there clearly exists $C > 0$ such that

$$\mathbb{P}\left[\int_0^{1/C} xH_{N,K}(dx) + \int_C^\infty xH_{N,K}(dx) \geq \epsilon\right] \leq \epsilon,$$

for large enough N. This implies that by uniformly approximating $f(x) = x$ through continous functions of the form

$$\tilde{f}(x) = \begin{cases} x, & x \in [1/C, C] \\ 0, & x \notin [1/2C, 2C] \end{cases} \quad \text{and} \quad \tilde{f}(x) \leq x, \ \forall x \in \mathbb{R}_+, \tag{1.68}$$

we have weak convergence of $Z_{a_N}(\beta)$ to $\int x H_K(dx) = \sum_{\mathbf{i}} \eta_{\mathbf{i}}$. This fact, together with the continuity of the mapping

$$\mathcal{M}_{mp}\left((\mathbb{R}^+)^{(2)} \times 2^I\right) \times (0,\infty) \to \mathcal{M}_{mp}\left((\mathbb{R}^+)^{(2)} \times 2^I\right)$$
$$\left(\sum_i \delta_{\{y_i;\, f_i\}},\, A\right) \mapsto \sum_i \delta_{\{y_i/A;\, f_i\}}$$

as well as Theorem 1.7 imply that $\Xi_{N,\beta}$ converges weakly to Ξ_β. It follows from the analysis carried out in [**5**] that the laws Ξ_β and $P_{x_K} \sqcap Q_{\mathbf{T},\mathbf{t}}$ coincide. This settles the proof of claim a).

The pure states. It is slightly more advantageous to know where exactly β lies (although it has no impact on the claim), so we suppose that $\beta \in (\beta_k, \beta_{k+1})$ for some $k \geq m$ *[since we do not specify k any further, we will cover the whole range $\beta > \beta_m$ at one fell swoop]*. It is more convenient to regard $\mathcal{G}^{(m)}_{\beta,N}$ as a marginal of $\mathcal{G}^{(k)}_{\beta,N}$. To this end, for $\sigma \in \Sigma_{N,A_m}$ we rewrite the points of the mth-marginal of the Gibbs measure as

$$\mathcal{G}^{(m)}_{\beta,N}(\sigma) = \sum_{\tau \in \Sigma_N : \tau_{A_m} = \sigma} \exp\left[\beta \widehat{X}_{\tau(1),\dots,\tau(k)} + \log \frac{Z_{\tau(1),\dots,\tau(k)}}{\mathbb{E}[\tau(1),\dots,\tau(k)]}\right] \Big/ Z_{a_N}(\beta).$$

By Proposition 1.12, to given $\epsilon > 0$ there exists $C > 0$ such that

$$\mathbb{P}\Big[\mathcal{G}^{(k)}_{\beta,N}\Big(\sigma \in \Sigma_N : \ \overline{X}_{\sigma(1),\dots,\sigma(l)} \in (-C, C)\ \forall l \leq k\Big) \geq 1 - \epsilon\Big] \geq 1 - \epsilon,$$

for large enough N. Then, as already pointed out at different occurences, there is also $\mathsf{N} = \mathsf{N}(\epsilon)$ such that $\mathbb{P}\big[\sharp\{\Sigma^C_{N,A_k}\} \geq \mathsf{N}\big] \leq \epsilon$. By Lemma 1.13, the fluctuations of these N associated random variables $\log Z_{\tau(1),\dots,\tau(k)}/\mathbb{E}[Z_{\tau(1),\dots,\tau(k)}]$ (the portion of the system in high-temperature) are negligible. This implies that in the weak limit of $\mathcal{G}^{(m)}_{\beta,N}$ coincide with the weak limits of the point process defined, for $\sigma \in \Sigma_{N,A_m}$, through

$$\widehat{\mathcal{G}}^{(m)}_{\beta,N}(\sigma) \stackrel{\text{def}}{=} \sum_{\substack{\tau \in \Sigma_{N,A_k}, \\ \tau_{A_m} = \sigma}} \frac{\exp\left[\beta \widehat{X}_{\tau(1),\dots,\tau(k)}\right]}{\widehat{Z}_m(\beta)}, \quad \widehat{Z}_m(\beta) \stackrel{\text{def}}{=} \sum_{\eta \in \Sigma_{N,A_k}} \exp\left[\beta \widehat{X}_{\eta(1),\dots,\eta(k)}\right].$$

We rewrite the points as $\widehat{\mathcal{G}}^{(m)}_{\beta,N}(\sigma) = \exp \beta\big[\widehat{X}_{\sigma(1),\dots,\sigma(m)} + U_{\sigma(1),\dots,\sigma(m)}\big] \big/ \widehat{Z}_m(\beta)$ where

$$U_{\sigma(1),\dots,\sigma(m)} = \frac{1}{\beta} \log \left\{ \sum_{\substack{\tau \in \Sigma_{N,A_k}, \\ \tau_{A_m} = \sigma}} \exp \beta \Big[\overline{X}_{\tau(1),\dots,\tau(m+1)} + \dots \overline{X}_{\tau(1),\dots,\tau(k)}\Big] \right\}.$$

Remark that to fixed $\sigma \in \Sigma_{N,A_m}$, $U_\sigma = U_{\sigma(1),\dots,\sigma(m)}$ is (up to a constant) the logarithm of the *partition function* of an irreducible hamiltonian in low temperature ($\beta > \beta_m$). A fixed realization $(\widehat{X}_{\sigma(1),\dots,\sigma(m)} + U_{\sigma(1),\dots,\sigma(m)}; \sigma \in \Sigma_{N,A_m})$ induces naturally an element of $\mathcal{M}_{mp}\big(\mathbb{R}^{(2)} \times 2^{A_m}\big)$. We denote by $\widehat{XU}_{N,m}$ its law. By Theorem 1.9, and the considerations in the proof of claim a) about the low temperature behavior of *partition functions* associated to an irreducible hamiltonian, we have that $\widehat{XU}_{N,m}$ converges weakly to the law $\widehat{XU}_m$ of the process on $\mathcal{M}_{mp}\big(\mathbb{R}^{(2)} \times 2^{A_m}\big)$ (with the possible marks being those from the restricted chain $\mathbf{T}^{(m)} = \{A_0, \dots, A_m\}$ only) induced by the collection of points given by $(u_{\mathbf{i}} + U_{\mathbf{i}}; \mathbf{i} \in \mathbb{N}^m)$ where

$$u_{\mathbf{i}} \stackrel{\text{def}}{=} u^1_{\mathbf{i}_1} + \cdots + u_{\mathbf{i}_m},\ U_{\mathbf{i}} \stackrel{\text{def}}{=} \frac{1}{\beta} \log \sum_{i_{m+1},\dots,i_k} \exp\Big[\beta\big(u^{m+1}_{\mathbf{i}_m, i_{m+1}} + \cdots + u^k_{\mathbf{i}_m, i_{m+1},\dots,i_k}\big)\Big].$$

For $l = 1, \dots, k$ and any multi-index $\mathbf{i}_{l-1}$ the point process $(u^l_{\mathbf{i}_{l-1}, i_l};\ i_l \in \mathbb{N})$ is poissonian with density $\mathcal{C}_l \beta_l \exp(-\beta_l t) dt$. The u^l are independent for different l and $(u^l_{\mathbf{i}_{l-1}, i_l};\ i_l \in \mathbb{N})$ are independent for different $\mathbf{i}_{l-1}$. An important observation is that to fixed $\mathbf{i}_{m-1}$ the PP $(u^m_{\mathbf{i}_{m-1}, i_m} + U_{\mathbf{i}_{m-1}, i_m};\ i_m \in \mathbb{N})$ is simply a shift by independent variables of a PPP, in which case it is easy to see that

$$\Big(u^m_{\mathbf{i}_{m-1}, i_m} + U_{\mathbf{i}_{m-1}, i_m} - const;\ i_m \in \mathbb{N}\Big) \stackrel{\text{(distr)}}{=} \Big(u^m_{\mathbf{i}_{m-1}, i_m};\ i_m \in \mathbb{N}\Big), \tag{1.69}$$

for some $const > 0$, cfr. [**4**, Prop. 8.7] and a straightforward generalization. By continuity under mappings, the process on $\mathcal{M}_{mp}\left((\mathbb{R}^+)^{(2)} \times 2^{A_m}\right)$ induced by the points $\Big(\exp \beta\big[\widehat{X}_{\sigma(1),\dots,\sigma(m)} + U_{\sigma(1),\dots,\sigma(m)} - const\big];\ \sigma \in \Sigma_{N,A_m}\Big)$ converges weakly to the process induced by the points $\big(\exp[\beta u_{\mathbf{i}}];\ \mathbf{i} \in \mathbb{N}^m\big)$. To get the weak limit of $\Xi^{(m)}_{\beta,N}$ it then suffices to prove that the normalization procedure commutes with the limit $N \to \infty$; this is done exactly as in case a); the proof of the Main Theorem is completed. $\square$

2. On a cavity field perturbation of the Random Energy Model

2.1. Introduction and outline

After years of intensive research and important advances ([**1**], [**13**], [**14**], [**19**]) the Parisi Picture [**16**] for mean field models of spin glasses still remains quite elusive. In fact, despite the spectacular proof by Guerra and Talagrand that the Replica Symmetry Breaking mechanism provides the correct answer on the level of the free energy of some notoriously difficult models as that of Sherrington and Kirkpatrick, there seems to be no idea around on how to address issues pertaining to the nature of the low temperature regime, such as the rôle of the Derrida-Ruelle hierarchical structures, the (related) ultrametricity, the law of the pure states, and the chaos problem, to quote a few.

In this work we gain some insights into these issues by swapping limits in the Cavity-RoSt framework of Aizenman, Sims and Starr; to wit, instead of taking the thermodynamical limit *first* and *then* perform a one spin perturbation of the Derrida-Ruelle structures, we perform a cavity field perturbation on the finite systems first, and only subsequently do we take the thermodynamical limit.

As a first step, we stick here to the simplest finite size counterpart of the Derrida-Ruelle structures, the Random Energy Model [**10**], REM for short. Despite its simplicity, this "REM+Cavity" shows a delicate phase transition where Replica Symmetry is broken and ultrametricity sets in. In the low temperature region massive pure states emerge, with law being given by the Poisson-Dirichlet distribution. Moreover, this model has chaotic behavior in the temperature. This all is the content of Section 2.3.

The scenario depicted above for the REM+Cavity steadily follows from a way more general approach which we develop at first place in Section 2.2. It allows to discuss in somewhat universal terms mean field models of spin glasses where certain dependency structures of REM-type can be recovered. The crucial ingredients are Large Deviations Principles on the level of the empirical measures and related Central Limit Theorems. Given the range of validity of these abstract theorems, we taste some flavour of universality of the Parisi Theory.

The proofs of the main results are collected in Section 2.4.

2.2. Mean field models of REM-type

Consider a double sequence $X_{\alpha,i}, \alpha, i \geq 1$ of i.i.d. random variables with a distribution μ, taking values in a Polish space $(S, \mathcal{S})$, and which are defined on a probability space

$(\Omega, \mathcal{F}, \mathbb{P})$. For $N \in \mathbb{N}$, let for every α, the empirical distributions be defined by

$$L_{N,\alpha} \stackrel{\text{def}}{=} \frac{1}{N}\sum_{i=1}^{N} \delta_{X_{\alpha,i}},$$

which takes values in $\mathcal{M}_1^+(S)$, the set of probability measures on $(S, \mathcal{S})$, which itself is a Polish space when equipped with the weak topology. Let $\Phi : \mathcal{M}_1^+ \to \mathbb{R}$ be a continuous function. We write

$$Z_N \stackrel{\text{def}}{=} 2^{-N}\sum_{\alpha=1}^{2^N} \exp\left[N\Phi(L_{N,\alpha})\right], \; f_N(\Phi,\mu) \stackrel{\text{def}}{=} \frac{1}{N}\log Z_N.$$

Theorem 2.1. *In the above setting we have:*

i) *The limit $f(\Phi,\mu) = \lim_{N\to\infty} f_N(\Phi,\mu)$ exists $\mathbb{P}-a.s.$*
ii) *Moreover, one has*

$$f(\Phi,\mu) = \sup\left\{\Phi(\nu) - H(\nu \mid \mu) : \; H(\nu \mid \mu) \le \log 2\right\},$$

where H is the usual relative entropy $H(\nu \mid \mu) \stackrel{def}{=} \int \log\left(\frac{d\nu}{d\mu}\right) d\nu$ if $\nu \ll \mu$ and $\log(d\nu/d\mu) \in L_1(\mu)$, and $= \infty$ otherwise.

We now specialize to linear functionals, thereby assuming that

$$Z_N = 2^{-N}\sum_{\alpha} \exp\left[\sum_{i=1}^{N} \phi(X_{\alpha,i})\right]$$

for reasonable $\phi : S \to \mathbb{R}$, in which case the functional from Theorem 2.1 reads $\Phi(\nu) = \int \phi(x)\nu(dx)$. By a slight abuse of notation, we write $f_N(\phi,\mu)$ for the free energy of the finite-size system, and $f(\phi,\mu)$ for its limit, which in virtue of Theorem 2.1 is given by

$$\sup\left\{\int \phi(x)\nu(dx) - H(\nu \mid \mu) : \; H(\nu \mid \mu) \le \log 2\right\}. \tag{2.1}$$

We shall refer to expression (2.1) as the Gibbs variational principle, GVP for short. In this setting, the GVP is strikingly simple to solve. We need some notation: for a distribution $\nu \in \mathcal{M}_1^+(S)$, and $h : S \to \mathbb{R}$ we write $\mathsf{E}_\nu[h] \stackrel{\text{def}}{=} \int h(x)\nu(dx)$. Moreover, for $m \in \mathbb{R}$, we introduce $\Gamma_\phi(m) \stackrel{\text{def}}{=} \log \mathsf{E}_\mu\left[e^{m\phi}\right]$.

Theorem 2.2. *There exists a unique $G \in \mathcal{M}_1^+(S)$ solving the Gibbs variational principle. It is characterized by the Radon-Nykodym derivative*

$$\frac{dG}{d\mu}(x) = \frac{e^{m_\star\phi(x)}}{\mathsf{E}_\mu\left[e^{m_\star\phi}\right]},$$

for $m_\star \in \mathbb{R}$ with the following property: if ϕ, μ are such that $\Gamma_\phi'(1) - \Gamma_\phi(1) \le \log 2$, then $m_\star = 1$, otherwise $m_\star \in (0,1)$ is solution to the following equation:

$$m\Gamma_\phi'(m) - \Gamma_\phi(m) = \log 2.$$

In terms of free energy of the system, Theorem 2.2 implies the following

Corollary 2.3. *$f_N(\phi,\mu)$ converges almost surely to the non-random functional:*

$$f(\phi,\mu) = \begin{cases} \Gamma_\phi(1), & \Gamma'_\phi(1) - \Gamma_\phi(1) \le \log 2, \\ \Gamma'_\phi(m_\star) - \log 2, & \textit{otherwise.} \end{cases} \tag{2.2}$$

where $m_\star \in (0,1)$ is the unique solution to the following equation:

$$m\Gamma'_\phi(m) - \Gamma_\phi(m) = \log 2. \tag{2.3}$$

We call the first case in (2.2) *high temperature*, and the second *low temperature*, whereas equation (2.3) is naturally an *entropy condition*.

In the case of linear functionals we introduce the finite-system Gibbs measure

$$\mathcal{G}_{\phi,N}(\alpha) \stackrel{\text{def}}{=} \exp\Big[\sum_{i=1}^N \phi(X_{\alpha,i})\Big] \Big/ Z_N, \quad \text{for } \alpha = 1,\dots,2^N, \tag{2.4}$$

In general, for a collection of random points $(\xi_i, i\in\mathbb{N})$ on the positive real line such that $\sum_{i\in\mathbb{N}} \xi_i < \infty$ almost surely, we define new points through the normalization $\overline{\xi}_i = \xi_i / \sum_j \xi_j$ and write $\mathcal{N}((\xi_i)) \stackrel{\text{def}}{=} (\overline{\xi}_i)$ for the normalization procedure. For a Poisson Point Process with intensity $r(t)$ we may also write $PPP(t \mapsto r(t))$.

Theorem 2.4. *Suppose that ϕ,μ are such that the system is in low temperature. Then the point process $\sum_\alpha \delta_{\mathcal{G}_{\phi,N}(\alpha)}$ converges weakly as $N\to\infty$ to a $\mathcal{N}(PPP(t\mapsto t^{-m_\star-1}))$, with $m_\star$ the unique solution to the entropy condition* (2.3)*.*

Remark that Theorem 2.4 accounts for some universality of the Derrida-Ruelle structures and the socalled Poisson-Dirichlet distribution, which naturally arise in the weak limits of the Gibbs measure associated to a REM-system in low temperature.

2.3. The REM+Cavity

We now come to the applications. Let again $N\in\mathbb{N}$. We set $\Sigma_N \stackrel{\text{def}}{=} \{1,\dots,2^N\}$ and consider on some probability space $(\Omega,\mathcal{F},\mathbb{P})$ a sequence $(X_\alpha, \alpha\in\Sigma_N)$ of independent, centered gaussians with variance N, as well as another independent sequence $(g_{\alpha,i}, \alpha \in \Sigma_N, i = 1,\dots,N)$ of standard gaussians. For $\alpha\in\Sigma_N, \sigma = (\sigma_1,\dots,\sigma_N) \in \{\pm 1\}^N$, we define the Hamiltonian of the REM+Cavity

$$H(\alpha,\sigma) \stackrel{\text{def}}{=} X_\alpha + \sum_{i=1}^N g_{\alpha,i}\sigma_i. \tag{2.5}$$

$H(\cdot,\cdot)$ is thus a gaussian field on $\Sigma_N \times \{\pm 1\}^N$ with covariance given by

$$\mathbb{E}[H(\alpha,\sigma)H(\alpha,\sigma')] = N\delta_{\alpha=\alpha'} + N\delta_{\alpha=\alpha'} q(\sigma,\sigma'),$$

where $q(\sigma,\sigma') \stackrel{\text{def}}{=} \frac{1}{N}\sum_{i=1}^N \sigma_i\sigma'_i$ is the usual overlap of the configurations σ,σ'. For $\beta\in\mathbb{R}$, the inverse of the temperature, we define the free energy of the REM+Cavity

$$f_N(\beta) \stackrel{\text{def}}{=} \frac{1}{N}\log\left[2^{-2N}\sum_{\alpha,\sigma}\exp\left(\beta H(\alpha,\sigma)\right)\right]. \tag{2.6}$$

(Here and henceforth, we denote by g a standard gaussian, and by E expectation with respect to it.)

Proposition 2.5. *The limit $f(\beta) = \lim_{N\to\infty} f_N(\beta)$ associated to the spin glass (2.5) exists $\mathbb{P}$-a.s. and coincides with $\lim_{N\to\infty} \mathbb{E}[f_N(\beta)]$. Moreover, we have*

$$f(\beta) = \begin{cases} \beta^2 & \beta \le \beta_\star, \\ \frac{\beta^2}{2} m_\star + \frac{\mathsf{E}[\cosh(\beta g)^{m_\star} \log\cosh(\beta g)]}{\mathsf{E}[\cosh(\beta g)^{m_\star}]} - \log 2 & \beta \ge \beta_\star, \end{cases}$$

with $\beta_\star$ being the unique positive solutions of the equation

$$\mathsf{E}[\cosh(\beta g)\log\cosh(\beta g)] = e^{\beta^2/2}\log 2, \tag{2.7}$$

and $m_\star = m_\star(\beta) \in (0,1)$ the unique solution of

$$\frac{\beta^2}{2} m^2 - \log \mathsf{E}[\cosh(\beta g)^m] + m\frac{\mathsf{E}[\cosh(\beta g)^m \log\cosh(\beta g)]}{\mathsf{E}[\cosh(\beta g)^m]} = \log 2. \tag{2.8}$$

According to the convention following Theorem 2.3, we call the region $\beta \le \beta_\star$ the high-temperature, and $\beta > \beta_\star$ the low temperature regime. Let us now introduce the Gibbs measure associated to the REM+Cavity:

$$\mathcal{G}_{\beta,N}(\alpha,\sigma) = \frac{\exp \beta H(\alpha,\sigma)}{\sum_{\alpha',\sigma'} \exp\beta H(\alpha',\sigma')}, \text{ for}(\alpha,\sigma) \in \Sigma_N \times \{\pm 1\}^N.$$

Our interest lies in the weak limit properties of the collection of points $(\mathcal{G}_{\beta,N}(\alpha,\sigma))$. It is not difficult to realize that even in low temperature no configuration gets a macroscopic weight when passing to the limit $N\to\infty$. To get something interesting we must lump together exponentially many configurations; to this end, let us call for $\alpha \in \Sigma_N$ the set $E_\alpha \stackrel{\text{def}}{=} \left\{(\alpha,\sigma) : \sigma \in \{\pm1\}^N\right\}$ a *pure state* and consider its mass under the Gibbs measure

$$\mathcal{G}_{\beta,N}(\alpha) \stackrel{\text{def}}{=} \sum_{\sigma\in\{\pm1\}^N} \mathcal{G}_{\beta,N}(\alpha,\sigma).$$

Proposition 2.6. *Let $\beta > \beta_\star$. Then the collection of points $\big(\mathcal{G}_{\beta,N}(\alpha); \alpha \in \Sigma_N\big)$ converges weakly to $\mathcal{N}(PPP(t \mapsto t^{-m_\star - 1}))$.*

We thus witness in the low-temperature regime of the REM+Cavity the emergence of massive pure states, with law (when reordered in non-increasing fashion) being given by the Poisson-Dirichlet distribution.

Despite the recent remarkable progress in the field of spin glasses of SK-type, the whole subject remains somewhat mysterious. For instance, the Parisi Theory predicts a very peculiar behavior of the overlap of two configurations belonging to the same pure state. In the most challenging models such as the original SK, the pure states are notoriously difficult to identify, and practically nothing is rigorously known to date as for their weak limits, the same being true for the microscopic properties of the system inside of a pure state. The REM+Cavity represents a simple ground where these issues can be addressed: suppose that $\beta > \beta_\star$ (low-temperature) and set

$$q_\star \stackrel{\text{def}}{=} \frac{\mathsf{E}\left[\tanh^2(\beta g)\exp\left(m_\star \log\cosh(\beta g)\right)\right]}{\mathsf{E}\left[\exp(m_\star \log\cosh(\beta g))\right]},$$

for g a standard gaussian and E denoting expectation with respect to it. Clearly, $m_\star$ is the unique solution to the entropy equation. For a function $F : \Sigma_N \times \Sigma_N \to \mathbb{R}$ on the replicated system we write

$$\left\langle F\right\rangle^{\otimes 2}_{\beta,N} \stackrel{\text{def}}{=} \sum_{(\alpha,\sigma;\alpha'\sigma')\in\Sigma_N\times\Sigma_N} F\big(\alpha,\sigma;\alpha',\sigma'\big)\mathcal{G}_{\beta,N}(\alpha,\sigma)\mathcal{G}_{\beta,N}(\alpha',\sigma').$$

Proposition 2.7 (Ultrametricity for the REM+Cavity)**.** *For $\beta > \beta_\star$,*

i) $\lim_{N\to\infty} \mathbb{E}\left[\left\langle \delta_{\alpha=\alpha'}\Big(q(\sigma,\sigma') - q_\star\Big)^2\right\rangle^{\otimes 2}_{\beta,N}\right] = 0.$

ii) $\lim_{N\to\infty} \mathbb{E}\left[\left\langle \delta_{\alpha\neq\alpha'} q(\sigma,\sigma')^2\right\rangle^{\otimes 2}_{\beta,N}\right] = 0,$

The above proves ultrametricity in the following simple terms: the overlap on the Ising-spins can take essentially two values, for large enough N. Inside of a pure state, the overlap of two configurations is constant, and equals $q_\star$ (case *i*), while the overlap of two configurations belonging to two different pure states (case *ii*) is essentially zero.

We conclude with a final remark pertaining to the chaos problem. In the physical literature, one says that a disordered system displays chaos as soon as the overlap associated to two systems at different temperatures can take one value only, the relevant configurations being therefore essentially unrelated. To date, the only models where these issues could be rigorously addressed are those of GREM-type; absence of chaos is however in these models quite evident, since (by construction), a change in the temperature does not affect the ordering of their energy levels. An interesting feature of the REM+Cavity is that, given the delicate β-dependence of the relevant energies, chaos is present. More precisely, consider the replicated space of configurations $(\Sigma_N \times \{\pm 1\}^N) \times (\Sigma_N \times \{\pm 1\}^N)$ endowed with the product Gibbs measure $\mathcal{G}_{\beta,N} \otimes \mathcal{G}_{\beta',N}$ given by the hamiltonians $\beta H(\alpha,\sigma)$ and $\beta' H(\alpha',\sigma')$ respectively. (This is nothing else then the product measure of two identical, i.e. with same disorder, copies at temperatures β and β' respectively). Let us also denote by $\langle\cdot\rangle^{\otimes 2}_{\beta,\beta',N}$ expectation with respect to $\mathcal{G}_{\beta,N} \otimes \mathcal{G}_{\beta',N}$.

Proposition 2.8 (Chaos in temperature for the REM+Cavity)**.** *Assume $\beta, \beta' > \beta_\star$ and $\beta \neq \beta'$.*

a) *Given $\varepsilon > 0$ there exists $\delta > 0$ such that*

$$\mathbb{P}\left[\mathcal{G}_{\beta,N} \otimes \mathcal{G}_{\beta',N}(\delta_{\alpha=\alpha'}) \geq e^{-\delta N}\right] \leq e^{-\delta N}.$$

b)

$$\lim_{N\to\infty} \mathbb{E}\left\langle \delta_{\alpha\neq\alpha'} q(\sigma,\sigma')^2\right\rangle^{\otimes 2}_{\beta,\beta',N} = 0.$$

2.4. Proofs of the main results

2.4.1. The free energy for spin glasses of REM-type. For the proof of Theorem 2.1, we still need a technical result. Given $A \subset \mathcal{M}_1^+(S)$, we set

$$M_N(A) \stackrel{\text{def}}{=} \sharp\left\{\alpha \le 2^N : L_{N,\alpha} \in A\right\}.$$

We also write $H(A)$ for $\inf_{\nu \in A} H(\nu \mid \mu)$.

Lemma 2.9. *Let $\nu \in \mathcal{M}_1^+(S)$, and V be an open neighborhood of ν. If $H(\nu \mid \mu) \le \log 2$, and $\varepsilon > 0$, then there exists an open neighborhood U of ν, $U \subset V$, and $\delta > 0$ such that for large enough N*

$$\mathbb{P}\Big[M_N(U) \le \exp\left[N(\log 2 - H(\nu \mid \mu) - \varepsilon)\right]\Big] \le e^{-N\delta}, \tag{2.9}$$

$$\mathbb{P}\Big[M_N(U) \ge \exp\left[N(\log 2 - H(\nu \mid \mu) + \varepsilon)\right]\Big] \le e^{-N\delta}. \tag{2.10}$$

If $H(\nu) > \log 2$, then there exist U and δ with

$$\mathbb{P}\Big[M_N(U) \ne 0\Big] \le e^{-N\delta}. \tag{2.11}$$

Proof. Let first $\nu \in \mathcal{M}_1^+(S)$ satisfy $H(\nu \mid \mu) \le \log 2$. The statement of the Lemma is trivial if $\nu = \mu$, so we assume $\nu \ne \mu$. Let $B_r(\nu) \subset \mathcal{M}_1^+(S)$ be the open ball of radius r and center ν, where we have equipped $\mathcal{M}_1^+(S)$ with one of the standard metrics, e.g. Prohorov's metric. Then $H(B_r(\nu)) = H(cl(B_r(\nu)))$, except for countably many r. Therefore we can find arbitrary small $\varepsilon_1 > 0$, and $U \stackrel{\text{def}}{=} B_r(\nu) \subset V$, such that $H(U) = H(cl(U)) = H(\nu \mid \mu) - \varepsilon_1$. From Sanov's Theorem, we have

$$\begin{aligned}
\mathbb{P}\Big[L_{N,\alpha} \in U\Big] &\ge \exp\Big[-N\Big(H(\nu \mid \mu) - \frac{5}{6}\varepsilon_1\Big)\Big],\\
\mathbb{P}\Big[L_{N,\alpha} \in cl(U)\Big] &\le \exp\Big[-N\Big(H(\nu \mid \mu) - \frac{7}{6}\varepsilon_1\Big)\Big],
\end{aligned} \tag{2.12}$$

for large enough N. Therefore

$$\begin{aligned}
&\mathbb{E}M_N(U) \ge \exp\Big[N\Big(\log 2 - H(\nu \mid \mu) + \frac{5}{6}\varepsilon_1\Big)\Big]\\
&\mathbb{E}M_N(U) \le \exp\Big[N\Big(\log 2 - H(\nu \mid \mu) + \frac{7}{6}\varepsilon_1\Big)\Big],\\
&\mathbb{E}\Big[M_N(U)^2\Big] \le \mathbb{E}[M_N(U)]^2 + \exp\Big[N\Big(\log 2 - H(\nu \mid \mu) + \frac{7}{6}\varepsilon_1\Big)\Big]\\
&\mathrm{var}_{\mathbb{P}}\Big(M_N(U)\Big) \le e^{-N\varepsilon_1/2}\Big(\mathbb{E}M_N(U)\Big)^2.
\end{aligned} \tag{2.13}$$

Hence,

$$\begin{aligned}
\mathbb{P}\Big[M_N(U) \le \Big(1 - e^{-N\varepsilon_1/8}\Big)\,\mathbb{E}M_N(U)\Big] &\le \exp\left[-N\varepsilon_1/4\right],\\
\mathbb{P}\Big[M_N(U) \ge \Big(1 + e^{-N\varepsilon_1/8}\Big)\,\mathbb{E}M_N(U)\Big] &\le \exp\left[-N\varepsilon_1/4\right].
\end{aligned} \tag{2.14}$$

Choosing ε_1 smaller than $\varepsilon/2$ and $\delta = \varepsilon_1/4$ proves the Lemma in this case. The case $H(\nu \mid \mu) > \log 2$ needs only a slight modification. In that case, there exists an open neighborhood $U \subset V$ of ν such that $\mathbb{P}\big[L_{N,\alpha} \in U\big]$ is exponentially small in N, with a decay rate which is bigger than $\log 2$. This proves the claim by the Markov inequality. $\square$

PROOF OF THEOREM 2.1. We first prove the lower bound. Let ν be any element in $\mathcal{M}_1^+(S)$ satisfying $H(\nu \mid \mu) \leq \log 2$. Let $\varepsilon > 0$. As Φ is continuous, we can choose an open neighborhood V of ν satisfying $\mid \Phi(\gamma) - \Phi(\nu) \mid \leq \varepsilon$ for $\gamma \in V$. Applying Lemma 2.9 we find a neighborhood U of ν in V satisfying (2.9). As

$$Z_N \geq 2^{-N} \exp\Big[N \inf_{\gamma \in U} \Phi(\gamma)\Big] M_N(U) \geq 2^{-N} e^{N(\Phi(\nu)-\varepsilon)} M_N(U),$$

we get from Lemma 2.9 that $\mathbb{P}$-a.s. one has eventually

$$\begin{aligned} Z_N &\geq 2^{-N} \exp\Big[N\left(\Phi(\nu) - \varepsilon\right)\Big] \exp\Big[N\left(\log 2 - H(\nu \mid \mu) - \varepsilon\right)\Big] \\ &\geq \exp\Big[N\left\{\Phi(\nu) - H(\nu \mid \mu) - 2\varepsilon\right\}\Big], \end{aligned} \tag{2.15}$$

and therefore

$$\liminf_{N\to\infty} \frac{1}{N} \log Z_N \geq \Phi(\nu) - H(\nu \mid \mu) \tag{2.16}$$

almost surely, for all ν with $H(\nu \mid \mu) \leq \log 2$. This proves the lower bound.

We now prove the upper bound. We use the well-known fact that there exists a compact set $K \subset \mathcal{M}_1^+(S)$ such that $\mathbb{P}\big[L_{N,\alpha} \notin K\big] \leq \exp\big[-N(\log 2 + 1)\big]$. Let D_N be the event

$$D_N \stackrel{\text{def}}{=} \bigcap_{\alpha=1}^{2^N} \{L_{N,\alpha} \in K\}.$$

Then

$$\mathbb{P}\left[D_N^c\right] \leq 2^N 2 \exp\Big[-N\left(\log 2 + 1\right)\Big], \tag{2.17}$$

and therefore

$$\mathbb{P}\left[\liminf_{N\to\infty} D_N\right] = 1.$$

Fix $\varepsilon > 0$. For any $\nu \in K$, we choose V_ν such that $\big|\Phi(\gamma) - \Phi(\nu)\big| \leq \varepsilon$ for $\gamma \in V_\nu$, and then $U_\nu \subset V_\nu$ according to Lemma 2.9. The U_ν cover K, and we choose a finite subcover, call it $U_{\nu_1}, \ldots, U_{\nu_m}$. Then, on $D \stackrel{\text{def}}{=} \liminf_N D_N$ we have, writing U_k instead of U_{ν_k},

$$\begin{aligned} Z_N &= 2^{-N} \sum_\alpha \exp\Big[N\Phi(L_{N,\alpha})\Big] \\ &= 2^{-N} \sum_{k=1}^m \sum_{\alpha : L_{N,\alpha} \in U_k} \exp\Big[N\Phi(L_{N,\alpha})\Big] \\ &\leq 2^{-N} \sum_{k=1}^m \exp\Big[N\left\{\Phi(\nu_k) + \varepsilon\right\}\Big] M_N(U_k) \\ &\leq \sum_{k : H(\nu_k|\mu) \leq \log 2} \exp\Big[N\left\{\Phi(\nu_k) - H(\nu_k \mid \mu) + 2\varepsilon\right\}\Big] \end{aligned} \tag{2.18}$$

outside an event which has probability at most $m \exp\left[-N \min_{j\leq m} \delta_j\right]$, where the δ_j corresponds to the U_j. From this estimate on gets that $\mathbb{P}$-a.s. one has

$$\limsup_{N\to\infty} \frac{1}{N} \log Z_N \leq \sup_{\nu : H(\nu|\mu)\leq \log 2} \left[\Phi(\nu) - H(\nu \mid \mu)\right],$$

which together with the lower bound settles the proof of Theorem 2.1. □

Proof of Theorem 2.2. We first restrict the analysis of the GVP to measures with Radon-Nykodym derivative being given by

$$\frac{d\nu}{d\mu} = \frac{\exp\left[m\phi(x)\right]}{Z(m)}, \text{ with } Z(m) = \mathsf{E}_\mu\left[e^{m\phi}\right]. \tag{2.19}$$

This evidently yields a lower bound to the GVP, which actually reads

$$\sup_{m\in\mathbb{R}} \left\{(1-m)\Gamma_\phi'(m) + \Gamma_\phi(m) : \quad m\Gamma_\phi'(m) - \Gamma_\phi(m) \leq \log 2\right\}. \tag{2.20}$$

We now claim that the target function $(1-m)\Gamma_\phi'(m)+\Gamma_\phi(m)$ is increasing on $m \in (-\infty, 1]$ and decreasing otherwise; in fact

$$\frac{d}{dm}\Big[(1-m)\Gamma_\phi'(m) + \Gamma_\phi(m)\Big] = (1-m)\Gamma_\phi''(m)$$

and convexity also implies that $\Gamma_\phi''(m) \geq 0$, $\forall m \in \mathbb{R}$. Thus, we can restrict the search for a maximizing $m \in \mathbb{R}$ in (2.20) to:

$$\sup_{m\in(-\infty,1]} \left\{(1-m)\Gamma_\phi'(m) + \Gamma_\phi(m) : \quad m\Gamma_\phi'(m) - \Gamma_\phi(m) \leq \log 2\right\} \tag{2.21}$$

But monotonicity also implies that the (global) maximum is attained in $m = 1$, as long as the side condition is satisfied:

$$\begin{aligned}\sup_{m\in(-\infty,1]} &\left\{(1-m)\Gamma_\phi'(m) + \Gamma_\phi(m) : \quad m\Gamma_\phi'(m) - \Gamma_\phi(m) \leq \log 2\right\} = \\ &= \Big((1-m)\Gamma_\phi'(m) + \Gamma_\phi(m)\Big)\Big|_{m=1} \\ &= \Gamma_\phi(1).\end{aligned} \tag{2.22}$$

In this case, the side condition reads $\Gamma_\phi'(1) - \Gamma_\phi(1) \leq \log 2$, the *high temperature*.

In the case where ϕ, μ are such that at $m = 1$ the constraint is not satisfied we first observe that the function giving rise to the side condition is also increasing, this time for any value of $m \geq 0$:

$$\frac{d}{dm}\Big(m\Gamma_\phi'(m) - \Gamma_\phi(m)\Big) = m\Gamma_\phi''(m).$$

Hence, monotonicity of both target and constraint function yields that the maximum is achieved at the largest possible value, which is the one satisfying:

$$m\Gamma_\phi'(m) - \Gamma_\phi(m) = \log 2 \tag{2.23}$$

To check existence and uniqueness of the solution to (2.23) we observe that

$$\Big[m\Gamma'_\phi(m) - \Gamma_\phi(m)\Big]\Big|_{m=0} = 0 < \log 2,$$

and that the *low temperature condition* ensures

$$\Big[m\Gamma'_\phi(m) - \Gamma_\phi(m)\Big]\Big|_{m=1} > \log 2.$$

The claim then follows in virtue of a simple result from Real Analysis, which also entails the property $m_\star \in (0,1)$.

We next claim that the extremal measures with Radon-Nykodym derivative as in (2.19), with either $m = 1$ or solution to (2.23), are maximimizers for the variational problem. In fact, we may certainly restrict the analysis to measures which are non-singular with respect to μ (if not, the relative entropy would run off to ∞). For convenience, call the Gibbs measures from the previous step G_m, that is:

$$\frac{dG_m}{d\mu} = e^{m\phi(x)}/Z(m),\ Z(m) = \mathsf{E}_\mu[e^{m\phi}].$$

Let us now suppose that there exists a measure $\nu \ll \mu$ satisfying the constraint

$$H(\nu \mid \mu) \leq \log 2 \tag{2.24}$$

for which

$$\mathsf{E}_\nu[\phi] - H(\nu \mid \mu) > \mathsf{E}_{G_m}[\phi] - H(G_m \mid \mu). \tag{2.25}$$

We compute the entropy of ν relative to G_m:

$$\begin{aligned} H(\nu \mid G_m) = \int d\nu \log\Big(\frac{d\nu}{dG_m}\Big) &= \int d\nu \log\Big(\frac{d\nu}{d\mu}\cdot\frac{d\mu}{dG_m}\Big) = \\ = H(\nu \mid \mu) - \mathsf{E}_\nu[\log \frac{dG_m}{d\mu}] &= H(\nu \mid \mu) - m\mathsf{E}_\nu[\phi] + \log Z(m) = \\ = H(\nu \mid \mu) - m\mathsf{E}_\nu[\phi] &+ m\mathsf{E}_{G_m}[\phi] - H(G_m \mid \mu) \end{aligned} \tag{2.26}$$

the last equality stemming from the definition of the G_m, according to which

$$H(G_m \mid \mu) = m\mathsf{E}_{G_m}[\phi] - \log Z(m).$$

Therefore,

$$
\begin{aligned}
H(\nu \mid G_m) &= H(\nu \mid \mu) - H(G_m \mid \mu) + m\Big\{ \mathsf{E}_{G_m}[\phi] - \mathsf{E}_\nu[\phi] \Big\} \\
&= H(\nu \mid \mu) - H(G_m \mid \mu) + \\
&\quad + m\Big\{ \Big[\mathsf{E}_{G_m}[\phi] - H(G_m \mid \mu)\Big] - \Big[\mathsf{E}_\nu[\phi] - H(\nu \mid \mu)\Big] \Big\} + \\
&\quad + mH(G_m \mid \mu) - mH(\nu \mid \mu) = \\
&= m\Big\{ \Big[\mathsf{E}_{G_m}[\phi] - H(G_m \mid \mu)\Big] - \Big[\mathsf{E}_\nu[\phi] - H(\nu \mid \mu)\Big] \Big\} + \\
&\quad + (m-1)\Big\{ H(G_m \mid \mu) - H(\nu \mid \mu) \Big\}
\end{aligned}
\tag{2.27}
$$

Let us abbreviate

$$
\Delta(G_m, \nu) \stackrel{\text{def}}{=} \Big\{ \Big[\mathsf{E}_{G_m}[\phi] - H(G_m \mid \mu)\Big] - \Big[\mathsf{E}_\nu[\phi] - H(\nu \mid \mu)\Big] \Big\}
$$

Remark that $\Delta(G_m, \nu) \leq 0$, for every ν fulfilling condition (2.25). We then have

$$
H(\nu \mid G_m) = m\Delta(G_m, \nu) + (m-1)\Big\{ H(G_m \mid \mu) - H(\nu \mid \mu) \Big\}
$$

Suppose we are in the high temperature regime: we know by now that there exists a measure G_m (and $m = 1$) satisfying the constraint, so that the second term on the r.h.s above drops out, and it remains $\Delta(G_m, \nu)$ which is by (2.25) negative.
On the other hand, in the low-temperature regime we can come up with the Gibbs measure associated to $m_\star$, the latter lying in $(0, 1)$: again, the first term on the r.h.s. is strictly less than 0, because of the assumption (2.25). The second term on the r.h.s. is also negative, since ν must satisfy the constraint (2.24), and the Gibbs measure associated to $m_\star$ is such that

$$
H(G_{m_\star} \mid \mu) = \log 2,
$$

and therefore:

$$
(m_\star - 1)\Big\{ H(G_{m_\star} \mid \mu) - H(\nu \mid \mu) \Big\} = (m_\star - 1)\Big\{ \log 2 - H(\nu \mid \mu) \Big\} \leq 0.
$$

In both high- and low-temperature regimes, the relative entropy $H(\nu \mid G_m)$ (either with $m = 1$ or $m = m_\star$) is negative, obviously a contradiction. This settles the proof of Theorem 2.2. $\square$

2.4.2. The Gibbs measure of spin glasses of REM-type. Let $m_\star$ be the unique solution to equation (2.3) and $G = G_{m_\star}$ the associated extremal measure. Set

$$\begin{aligned} &\mathsf{E}_G(\phi) \stackrel{\text{def}}{=} \int \phi(x)G(dx),\ \mathsf{V}(\phi) \stackrel{\text{def}}{=} \int \phi(x)^2 G(dx) - \mathsf{E}_G(\phi)^2, \\ &\qquad \text{as well as} \quad a_N \stackrel{\text{def}}{=} \mathsf{E}(\phi)N + \omega(N),\ \omega(N) \stackrel{\text{def}}{=} -\frac{1}{m_\star}\log\sqrt{2\pi\mathsf{V}(\phi)N}. \end{aligned} \tag{2.28}$$

For $\alpha \in \{1,\dots,2^N\}$ let us also abbreviate $H_N(\alpha) \stackrel{\text{def}}{=} \sum_{i=1}^N \phi(X_{\alpha,i})$.

Lemma 2.10. *With the above notations, for $N \to \infty$:*

$$\mathbb{P}\Big[\sum_{i=1}^N \phi(X_{1,i}) - a_N \geq t\Big] = 2^{-N}\left\{\frac{1}{m_\star}\exp\big(-m_\star t\big) + o(1)\right\}. \tag{2.29}$$

Proof. It essentially follows from the CLT for iid random variables. Standard reference is the monograph [**17**]. Similar computations have been exploited recently in [**9**] for the asymptotics of sums of heavy-tailed random variables.

Consider a sequence of iid (real valued) random variables $(\widetilde{X}_{1,i}; i = 1,\dots N)$ defined on some additional probability space $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \mathbb{Q})$, distributed according to $G(dx)$. We also set $G_N(-\infty, y] \stackrel{\text{def}}{=} \mathbb{Q}\Big[\sum_{i\leq N}\phi(\widetilde{X}_{1,i}) - \mathsf{E}_G(\phi) \leq y\Big]$, $\psi(x)$ stand for the density of a standard gaussian and $\Psi(x) \stackrel{\text{def}}{=} \int_{[x,\infty)}\psi(y)dy$. (Remark that $\mathsf{E}_G[\phi]$ is nothing more than the mean value of the random variable $\phi(\widetilde{X}_{1,1})$ under $\mathbb{Q}$.)

By change of measure and integration by parts, it holds:

$$\begin{aligned} &\mathbb{P}\Big[\sum_{i=1}^N \phi(X_{1,i}) - a_N \geq t\Big] = \\ &= \exp N\Big\{\log \mathsf{E}[e^{m_\star\phi}] - m_\star \mathsf{E}_G(\phi)\Big\}\int_{t+\omega(N)} e^{-m_\star y} G_N(dy) = \\ &= \exp N\Big\{\log \mathsf{E}[e^{m_\star\phi}] - m_\star \mathsf{E}_G(\phi)\Big\} \times \Bigg\{ -\exp\big[-m_\star y\big]G_N[y,\infty)\Big|_{-\infty}^{t+\omega(N)} \\ &\qquad\qquad - m_\star\int_{t+\omega(N)}^\infty \exp\Big[-m_\star y\Big] G_N[y,\infty)dy\Bigg\}. \end{aligned} \tag{2.30}$$

By [**17**, Th. 5.22], the following uniform bound holds

$$\begin{aligned} G_N[y,\infty) &= \mathbb{Q}\left[\frac{\sum_{i=1}^N \phi(\widetilde{X}_{1,i}) - \mathsf{E}_G(\phi)N}{\sqrt{N\mathsf{V}(\phi)}} \geq \frac{y}{\sqrt{N\mathsf{V}(\phi)}}\right] = \\ &= \quad \Psi\Big(y/\sqrt{N\mathsf{V}(\phi)}\Big) - \frac{const}{\sqrt{N\mathsf{V}(\phi)}}\Big(1 - \frac{y^2}{N\mathsf{V}(\phi)}\Big)\psi\big(\frac{y}{\sqrt{N\mathsf{V}(\phi)}}\big) + o(N^{-1/2}). \end{aligned} \tag{2.31}$$

Recall also that by the entropy condition, $\log \mathsf{E}[e^{m_\star \phi}] - m_\star \mathsf{E}_G(\phi) = -\log 2$. This, and (2.31) inserted in (2.30) yields

$$\mathbb{P}\Big[\sum_{i=1}^N \phi(X_{1,i}) - a_N \geq t\Big] = 2^{-N}\Big\{(I) + (II) + (III)\Big\},$$

where

$$\begin{aligned}
(I) &\stackrel{\text{def}}{=} -\exp\big[-m_\star y\big]\Psi\big(y/\sqrt{N\mathsf{V}(\phi)}\big)\Big|_{t+\omega(N)}^{\infty} - m_\star \int_{t+\omega(N)}^{\infty} e^{-m_\star y}\Psi\Big(y/\sqrt{N\mathsf{V}(\phi)}\Big)dy = \\
&= \exp\Big[\frac{m_\star^2}{2}N\mathsf{V}(\phi)\Big]\int_t^\infty \exp\left[-\frac{(z - m_\star N\mathsf{V}(\phi) + \omega(N))^2}{2N\mathsf{V}(\phi)}\right]\frac{dz}{\sqrt{2\pi N\mathsf{V}(\phi)}} = \\
&= \int_t^\infty e^{-m_\star y + o(1)}dy,
\end{aligned}$$

uniformly. As for the second term, we have:

$$\begin{aligned}
(II) &\stackrel{\text{def}}{=} \exp\big[-m_\star y\big]\frac{const}{\sqrt{N\mathsf{V}(\phi)}}\Big[1 - \frac{y^2}{N\mathsf{V}(\phi)}\Big]\psi\Big(y/\sqrt{N\mathsf{V}(\phi)}\Big)\Big|_{t+\omega(N)}^{\infty} + \\
&\qquad + m_\star \int_{t+\omega(N)}^{\infty} e^{-m_\star y}\frac{const}{\sqrt{N\mathsf{V}(\phi)}}\Big[1 - \frac{y^2}{N\mathsf{V}(\phi)}\Big]\psi(y/\sqrt{N\mathsf{V}(\phi)})dy \\
&= \int_{t+\omega(N)}^{\infty} e^{-m_\star y}\frac{1}{\sqrt{N\mathsf{V}(\phi)}}\psi(\frac{y}{\sqrt{N\mathsf{V}(\phi)}})dy + \\
&\qquad + \frac{const}{\sqrt{N\mathsf{V}(\phi)}}\int_{t+\omega(N)}^{\infty} e^{-m_\star y}\frac{d}{dy}\left[\Big(1 - \frac{y^2}{N\mathsf{V}(\phi)}\Big)\exp\big[-\frac{y^2}{2N\mathsf{V}(\phi)}\big]\right]dy \\
&= const \int_{\frac{t+\omega(N)}{\sqrt{N\mathsf{V}(\phi)}}}^{\infty} (z^3 - 3z)e^{-z^2/2}e^{-mz\sqrt{N\mathsf{V}(\phi)}}\frac{dz}{\sqrt{N\mathsf{V}(\phi)}} = O(N^{-1/2}).
\end{aligned}$$

Finally,

$$(III) \stackrel{\text{def}}{=} -e^{-m_\star y}o(N^{-1/2})\Big|_{t+\omega(N)}^{\infty} - m_\star \int_{t+\omega(N)}^{\infty} e^{-m_\star y}o(N^{-1/2})dy,$$

which by definition of $\omega(N)$ reads then

$$= \exp\Big[-m_\star\big(t + \omega(N)\big)\Big]o(N^{-1/2}) = N^{1/2}o(N^{-1/2}) = o(1).$$

Altogether we have

$$\mathbb{P}\Big[\sum_{i=1}^N \phi(X_{1,i}) - a_N \geq t\Big] = 2^{-N}\Big[\int_t^\infty e^{-m_\star y + o(1)}dy + o(1)\Big].$$

which settles the Lemma. □

Proposition 2.11. *Within the above setting:*

a) *the Point Process $\big(H_N(\alpha) - a_N; \alpha \in \Sigma_N\big)$ converges weakly to a Poisson Point Process with density $e^{-m_\star t}dt$;*

b) *the Point process* $\big(\exp(H_N(\alpha) - a_N); \alpha \in \Sigma_N\big)$ *converges weakly to a PPP of density on the positive axis* $t^{-m_\star - 1}dt$.

PROOF. For the first claim, we shall exploit the equivalence of weak convergence and convergence of Laplace functionals. For a continuous function with compact support $F \in C_o(\mathbb{R})$ we have:

$$
\begin{aligned}
&\mathbb{E}\Big[\exp\Big(-\sum_\alpha F(H_N(\alpha) - a_N)\Big)\Big] = \\
&\quad = \mathbb{E}\Big[\exp\Big(-F(H_N(1) - a_N)\Big)\Big]^{2^N} = \\
&\quad = \Big\{1 - \mathbb{E}\Big[1 - \exp\Big(-F(H_N(1) - a_N)\Big)\Big]\Big\}^{2^N} = \\
&\quad = \Bigg\{1 - \int \Big(1 - e^{-F(t)}\Big)\mathbb{P}\Big[\sum_i \phi(X_{1,i}) - a_N \in dt\Big]\Bigg\}^{2^N},
\end{aligned}
\tag{2.32}
$$

which by Lemma 2.10 converges to $\exp\Big[-\int \big(1 - e^{-F(t)}\big)e^{-m_\star t}dt\Big]$, settling the proof of part a). Part b) follows by well known properties of PPP. $\square$

Since the law of $\big(\mathcal{G}_{N,\phi}(\alpha); \alpha \in \Sigma_N\big)$ coincides with that of

$$
\Big(\exp\big[H_N(\alpha) - a_N\big] \Big/ \sum_{\alpha'} \exp\big[H_N(\alpha') - a_N\big]; \alpha' \in \Sigma_N\Big),
$$

to prove Theorem 2.4 it suffices to prove that in low temperature the normalization commutes with taking the $N \to \infty$ limit. For this, we have the following:

Lemma 2.12. *Suppose ϕ is such that the system is in low temperature, and let $\varepsilon > 0$. There exists $C > 0$ such that*

$$
\mathbb{P}\Big[\sum_\alpha \exp\big[H_N(\alpha) - a_N\big] 1_{H_N(\alpha) - a_N \notin [-C, C]} \geq \varepsilon\Big] \leq \varepsilon,
$$

for large enough N.

PROOF. We clearly have that (for large enough N)

$$
\begin{aligned}
&\mathbb{P}\Big[\sum_\alpha \exp\big[H_N(\alpha) - a_N\big] 1_{H_N(\alpha) - a_N \geq C} \geq \varepsilon\Big] \\
&\qquad \leq \mathbb{P}\Big[\exists \alpha \in \Sigma_N : H_N(\alpha) - a_N \geq C\Big] \\
&\qquad \leq 2^N \mathbb{P}\Big[H_N(1) - a_N \geq C\big] \leq \exp\big[-m_\star C + const\big],
\end{aligned}
\tag{2.33}
$$

the last inequality by Lemma 2.10. This term can be made arbitrarily small by choosing C large enough in the positive. So, it remains to prove that we can find C large enough in the positive such that

$$
\mathbb{P}\Big[\sum_{\alpha \in \Sigma_N} \exp\big[H_N(\alpha) - a_N\big] 1_{H_N(\alpha) - a_N \leq -C} \geq \varepsilon\Big] \leq \frac{\varepsilon}{2}.
$$

To see the last inequality, remark that the l.h.s. is clearly bounded above by

$$\frac{2^N}{\varepsilon}\mathbb{E}\Big[\exp\Big\{H_N(1)-a_N\Big\}1_{H_N(1)-a_N\leq -C}\Big].$$

As for a useful bound to the expectation, we proceed along the lines of Lemma 2.10, by first change of measure and then partial integration:

$$\begin{aligned}
&\mathbb{E}\Big[\exp\Big\{H_N(1)-a_N\Big\}1_{H_N(1)-a_N\leq -C}\Big]\\
&= e^{N\log \mathsf{E}[\exp m\phi]-N\mathsf{E}_G(\phi)}e^{-\omega(N)}\int_{-\infty}^{-C+\omega(N)}\exp\big[(1-m_\star)y\big]G_N(dy)\\
&= 2^{-N}e^{-\omega(N)}\Bigg\{\int_{-\infty}^{-C+\omega(N)}e^{(1-m_\star)y}\psi(y/\sqrt{N\mathsf{V}(\phi)})\frac{dy}{\sqrt{N\mathsf{V}(\phi)}}+\\
&\qquad +\frac{const}{\sqrt{N\mathsf{V}(\phi)}}\int_{-\infty}^{-C+\omega(N)}e^{(1-m_\star)y}\frac{d}{dy}\left[\Big(1-\frac{y^2}{N\mathsf{V}(\phi)}\Big)\psi(y/\sqrt{N\mathsf{V}(\phi)})\right]dy+\\
&\qquad + e^{(1-m_\star)y}o(N^{-1/2})\Big|_{-\infty}^{-C+\omega(N)}+\int_{-\infty}^{-C+\omega(N)}e^{(1-m_\star)y}o(N^{-1/2})dy\Bigg\}
\end{aligned}\tag{2.34}$$

The crucial point here is that $m_\star\in(0,1)$ so that all the integrals above exist (most notably, the one involving the $o(N^{-1/2})$-term). It is easy to see that the r.h.s of (2.34) is of order

$$2^{-N}\exp\Big[-\underbrace{(1-m_\star)}_{>0}C+o(1)\Big]$$

for N large enough. This can be made as small as needed by choosing C large enough. □

Proof of Theorem 2.4. We denote by M the space of Radon measures on $(0,\infty)$ endowed with the vague topology. By $\mathcal{H}_N$ we denote the point process associated to the collection of points $(\exp[H_N(\alpha)-a_N],\alpha\in\Sigma_N)$ and $\mathcal{H}$ be its vague limit. We choose a continuous function $h:\mathbb{R}_+\to\mathbb{R}_+$ with $h(x)=x$ for $x\in[1/C,C]$, $h(x)\leq x\ \forall x$ and $h(x)=0$ for $x\notin[1/2C,2C]$. Then $\int h\,d\mathcal{H}_N$ converges weakly to $\int h\,d\mathcal{H}$ by continuity of the mapping $M\ni\Xi\to\int hd\Xi$. By Lemma 2.12, to $\varepsilon>0$ we can find $C>0$ large enough such that

$$\mathbb{P}\left[\int_0^{1/C}xd\mathcal{H}_N+\int_C^\infty xd\mathcal{H}_N\geq\varepsilon\right]\leq\varepsilon,$$

uniformly in N, from which we see by approximation that $\int_0^\infty xd\mathcal{H}_N$ converges weakly to $\int_0^\infty xd\mathcal{H}$. This also implies that $(\mathcal{H}_N,\int_0^\infty xd\mathcal{H}_N)$ converges weakly towards $(\mathcal{H},\int_0^\infty x\,d\mathcal{H})$. Theorem 2.4 then clearly follows from the continuity of the mapping $M\times(0,\infty)\to M$ defined through $(\Xi,a)\mapsto\Xi\theta_a^{-1}$ with $\theta_a:(0,\infty)\to(0,\infty)$ and $\theta_a(x)\stackrel{\text{def}}{=}x/a$. □

2.4.3. The free energy of the REM+Cavity. We denote by g a standard centered gaussian and by E expectation with respect to it.

Proof of Proposition 2.5. The fact that the free energy $f_N(\beta)$ is in the limit self-averaging, i.e. that $\lim_N f_N(\beta)$ (if the limit exists) coincides with $\lim_N \mathbb{E}[f_N(\beta)]$ is a simple consequence of the Gaussian Concentration Phenomenon, cfr. e.g. [**15**]. This seems to be a typical feature of Mean Field Models for Spin Glasses.

As for the limiting free energy, we write the $f_N(\beta)$ in such a way that the abstract theorems from Section 2.2 apply: firstly, performing the trace over the Ising-spins we have:

$$f_N(\beta) = \frac{1}{N}\log 2^{-N}\sum_\alpha \exp\left[\beta X_\alpha + \sum_i \log\cosh(\beta g_{\alpha,i})\right] \tag{2.35}$$

We may then regard the X_α's as the sum of N independent standard gaussians $X_{\alpha,i}$, i.e. $X_\alpha = \sum_{i=1}^N X_{\alpha,i}$ (such a decomposition clearly does not modify the statistics of the system) so that we get

$$f_N(\beta) = \frac{1}{N}\log 2^{-N}\sum_\alpha \exp\left[N\Phi(L_{N,\alpha})\right]$$

with $\Phi : \mathcal{M}_1^+(\mathbb{R}^2) \to \mathbb{R}$, $\nu \mapsto \int \phi(x_1,x_2)\nu(dx_1,dx_2)$ and $\phi(x_1,x_2) = \beta x_1 + \log\cosh(\beta x_2)$ and $L_{N,\alpha} \stackrel{\text{def}}{=} \frac{1}{N}\sum_{i=1}^N \delta_{X_{\alpha,i},g_{\alpha,i}}$.

Theorem 2.3 does not yet apply, as ϕ is not of bounded support. However, only a minor modification of the argument is required: We introduce, for $L \in \mathbb{R}_+$,

$$\phi_L(x_1,x_2) \stackrel{\text{def}}{=} \begin{cases} \phi(x_1,x_2) & \text{if} \quad (x_1,x_2) \in [-L,L]^2, \\ 0 & \text{otherwise.} \end{cases}$$

Set also $\Phi_L(\nu) \stackrel{\text{def}}{=} \int \phi_L(x_1,x_2)\nu(dx_1,dx_2)$ and $\widehat{\Phi}_L(\nu) = \int(\phi-\phi_L)d\nu$.

$$f_N(\beta)] = \frac{1}{N}\log 2^{-N}\sum_\alpha \exp\left(N\Phi_L(L_{N,\alpha} + N\widehat{\Phi}(L_{N,\alpha}))\right)$$

Since $\lim_{x\to\pm\infty}(1/x)\log\cosh(\beta x) = \beta$, by Jensen's inequality and the usual gaussian estimates, it is evident that

$$\limsup_{L\to+\infty}\limsup_{N\to\infty}\mathbb{E}\left[\frac{1}{N}\log\sum_\alpha 2^{-N}\exp\left(N\widehat{\Phi}_L(L_{N,\alpha})\right)\right] = -\infty,$$

thus

$$\lim_{N\to\infty}\mathbb{E}[f_N(\beta)] = \lim_{L\to\infty}\lim_{N\to\infty}\mathbb{E}\left[\frac{1}{N}\log\sum_\alpha 2^{-N}\exp\left(\Phi_L(L_{N,\alpha})\right)\right]. \tag{2.36}$$

Theorem 2.3 now yields, to fixed L, the $\mathbb{P}$-a.s. asymptotics of

$$f(\beta,L) \stackrel{\text{def}}{=} \lim_N \frac{1}{N}\log\sum_\alpha 2^{-N}\exp\left(\Phi_L(L_{N,\alpha})\right).$$

(Again by concentration of measure, it coincides with the $L_1(\mathbb{P}(d\omega))$ limit.) Then one can easily remove the cutoff, getting that $\lim_{L\to\infty} f(\beta, L) = f(\beta)$; in fact, the critical temperature $\beta_\star$ separating the high and low temperature regimes is related to the threshold

$$\Gamma'_\phi(1) - \Gamma_\phi(1) \le \log 2 \Longleftrightarrow \mathsf{E}\Big[\cosh(\beta g)\log\cosh(\beta g)\Big] \le e^{\beta^2/2}\log 2.$$

The claim on $\beta_\star$ now follows in virtue of Lemma 2.13 below. This concludes the proof of the Proposition. □

Lemma 2.13. a) *For every $\beta \in \mathbb{R}$ it holds:*

$$\mathsf{E}[\log\cosh(\beta g)] \le e^{-\beta^2/2}\mathsf{E}[\cosh(\beta g)\log\cosh(\beta g)].$$

b) *There exists a unique $\beta_\star > 0$ such that:*

$$\mathsf{E}[\cosh(\beta_\star g)\log\cosh(\beta_\star g)] = e^{\beta_\star^2/2}\log 2. \tag{2.37}$$

For $\beta \le \beta_\star$

$$\mathsf{E}[\cosh(\beta g)\log\cosh(\beta g)] \le e^{\beta^2/2}\log 2$$

whereas for $\beta > \beta_\star$

$$\mathsf{E}[\cosh(\beta g)\log\cosh(\beta g)] > e^{\beta^2/2}\log 2.$$

Proof. The m-derivative of the function $\Gamma_\beta(m) \stackrel{\text{def}}{=} \log\mathsf{E}\big[e^{m\log\cosh(\beta g)}\big]$ reads

$$\Gamma'_\beta(m) = \frac{\mathsf{E}[\cosh(\beta g)^m\log\cosh(\beta g)]}{\mathsf{E}[\cosh(\beta g)^m]}$$

Since $\Gamma_\beta(\cdot)$ is a convex function, its derivative is increasing, and thus $\Gamma'_\beta(0) \le \Gamma'_\beta(1)$. This yields claim a). As for claim b), let $H(\beta) \stackrel{\text{def}}{=} e^{-\beta^2/2}\mathsf{E}[\log\cosh(\beta g)\cosh(\beta g)]$. Clearly, $H(0) = 0$ and by a)

$$H(\beta) \ge \mathsf{E}[\log\cosh(\beta g)]$$

implying that H is not bounded above. Exploiting integration by parts for gaussian r.v.'s it holds:

$$\begin{aligned}
\frac{d}{d\beta}H(\beta) &= -\beta e^{-\beta^2/2}\mathsf{E}[\cosh(\beta g)\log\cosh(\beta g)]+\\
&\qquad + e^{-\beta^2/2}\mathsf{E}[g\sinh(\beta g)] + e^{-\beta^2/2}\mathsf{E}[g\log\cosh(\beta g)\sinh(\beta g)] =\\
&= \beta e^{-\beta^2/2}\Bigg\{ - \mathsf{E}[\cosh(\beta g)\log\cosh(\beta g)]+\\
&\qquad + \mathsf{E}[\cosh(\beta g)] + \mathsf{E}[\sinh^2(\beta g)\cosh(\beta g)^{-1}] + \mathsf{E}[\log\cosh(\beta g)\cosh(\beta g)]\Bigg\} =\\
&= \beta e^{-\beta^2/2}\Bigg\{\mathsf{E}[\cosh(\beta g)] + \mathsf{E}[\sinh^2(\beta g)\cosh(\beta g)^{-1}]\Bigg\} \ge 0
\end{aligned}$$

Hence, H is also increasing. □

2.4.4. The Gibbs measure of the REM+Cavity.

PROOF OF PROPOSITION 2.6. Performing the trace over the Ising spins, the Gibbs weight of the pure state $\alpha \in \{1, \ldots, 2^N\}$ in the REM+Cavity reads

$$\mathcal{G}_{\beta,N}(\alpha) = \exp\left[\beta X_\alpha + \sum_{i=1}^N \log\cosh(\beta g_{\alpha,i})\right] \Big/ \sum_{\alpha'} \exp\left[\beta X_{\alpha'} + \sum_{i=1}^N \log\cosh(\beta g_{\alpha',i}).\right]$$

As in the proof of the Proposition 2.5, we may then replace the X_α with $\sum_{i=1}^N X_{\alpha,i}$ for a double sequence of independent standard gaussians $X_{\alpha,i}$ (this evidently does not affect the distribution of the Gibbs weights). Theorem 2.4 then clearly applies. $\square$

We next recall some remarkable properties of the Derrida-Ruelle processes in the case of 1-step Replica Symmetry Breaking (the proof can be found for instance in [**19**, Theorem 6.4.5]).

Lemma 2.14. *Assume that the sequence of weights (v_α) is distributed like $PPP(t^{-m_\star-1}dt)$ and is independent of the sequence (U_α, V_α). Assume also that $V \geq 1, \mathsf{E}[U^2], \mathsf{E}[V^2] < \infty$. Then the following formulas hold:*

1. $\mathsf{E}\left[\frac{\sum_\alpha v_\alpha U_\alpha}{\sum_\alpha v_\alpha V_\alpha}\right] = \frac{\mathsf{E}\left[UV^{m_\star-1}\right]}{\mathsf{E}[V^{m_\star}]}$
2. $\mathsf{E}\left[\frac{\sum_{\alpha\neq\beta} v_\alpha v_\beta U_\alpha U_\beta}{\left(\sum_\alpha v_\alpha V_\alpha\right)^2}\right] = m_\star \left(\frac{\mathsf{E}\left[UV^{m_\star-1}\right]}{\mathsf{E}[V^{m_\star}]}\right)^2$
3. $\mathsf{E}\left[\frac{\sum_\alpha v_\alpha^2 U_\alpha^2}{\left(\sum_\alpha v_\alpha V_\alpha\right)^2}\right] = (1-m_\star)\frac{\mathsf{E}\left[U^2V^{m_\star-2}\right]}{\mathsf{E}[V^{m_\star}]}$

With these formulae we can tackle the issue of ultrametricity in the REM+Cavity:

PROOF OF PROPOSITION 2.7. Let

$$w_\alpha^{(1,2)} \stackrel{\text{def}}{=} \exp\left(\beta X_\alpha + \sum_{i=3}^N \log\cosh(\beta g_{\alpha,i}) - a_N\right)$$

stand for the "Boltzmann weight" of the pure state α with a cavity in the sites $i = 1, 2$, and constant a_N being given by (2.28), obviously specialized to the setting of the REM+Cavity. Remark that these weights somewhat differ from the original ones without the cavity, but only a fairly trivial modification to the considerations in the proof of Proposition 2.11 is needed to get the weak convergence of $(w_\alpha^{(1,2)})$ towards a collection (v_α) distributed according to a PPP of density $t^{-m_\star-1}dt$ on $\mathbb{R}_+$.

We also write

$$w_\alpha^{(1)} \stackrel{\text{def}}{=} \exp\left(\beta X_\alpha + \sum_{i=2}^N \log\cosh(\beta g_{\alpha,i}) - a_N\right).$$

Again, $\left(w_\alpha^{(1)}\right)$ evidently weakly converges to a PPP with density $t^{-m_\star-1}dt$ on $\mathbb{R}_+$.

As for the proof of claim $i)$ in Proposition 2.7: expanding the quadratic terms, by symmetry and obvious bounds one easily sees that

$$\begin{aligned}&\mathbb{E}\left[\left\langle \delta_{\alpha=\alpha'}\Big(q(\sigma,\sigma')-q_\star\Big)^2\right\rangle_{\beta,N}^{\otimes 2}\right]=\\&=\{1+O(1/N)\}\,\mathbb{E}\left[\left\langle \delta_{\alpha=\alpha'}\sigma_1\sigma_2\sigma_1'\sigma_2'\right\rangle_{\beta,N}^{\otimes 2}\right]-2q_\star\mathbb{E}\left[\left\langle \delta_{\alpha=\alpha'}\sigma_1\sigma_1'\right\rangle_{\beta,N}^{\otimes 2}\right]+q_\star^2\mathbb{E}\left\langle \delta_{\alpha,\alpha'}\right\rangle_{\beta,N}^{\otimes}.\end{aligned}\tag{2.38}$$

Performing the trace over the Ising spins separately (first in the numerator and then in the denominator) we get

$$\begin{aligned}(2.38)=\{1+O(1/N)\}\,\mathbb{E}&\left[\frac{\sum_\alpha\{\sinh(\beta g_{\alpha,1})\sinh(\beta g_{\alpha,2})\}^2\left(w_\alpha^{(1,2)}\right)^2}{\left(\sum_\alpha\cosh(\beta g_{\alpha,1})\cosh(\beta g_{\alpha,2})w_\alpha^{(1,2)}\right)^2}\right]+\\&-2q_\star\mathbb{E}\left[\frac{\sum_\alpha\sinh(\beta g_{\alpha,1})^2\left(w_\alpha^{(1)}\right)^2}{\left(\sum_\alpha\cosh(\beta g_{\alpha,1})w_\alpha^{(1)}\right)^2}\right]+q_\star^2\mathbb{E}\left[\langle\delta_{\alpha=\alpha'}\rangle_{\beta,N}^{\otimes 2}\right]=\\&=(I)+(II)+(III).\end{aligned}\tag{2.39}$$

As for the last term on the r.h.s. above, by Proposition 2.6, $(\mathcal{G}_{\beta,N}(\alpha))$ converges to a $\mathcal{N}(PPP(t^{-m_\star-1}dt))$, from which one easily deduces that

$$\mathbb{E}\left[\langle\delta_{\alpha=\alpha'}\rangle_{\beta,N}^{\otimes 2}\right]=\mathbb{E}\left[\sum_\alpha\mathcal{G}_{\beta,N}(\alpha)^2\right]\to 1-m_\star,\ \text{as } N\to\infty.\tag{2.40}$$

As for the term before last, by Proposition 2.11, together with similar considerations as in the proof of Theorem 2.4 and by Lemma 2.14 (formula 3) we have

$$\lim_{N\to\infty}\mathbb{E}\left[\frac{\sum_\alpha\sinh(\beta g_{\alpha,1})^2\left(w_\alpha^{(1)}\right)^2}{\left(\sum_\alpha\cosh(\beta g_{\alpha,1})w_\alpha^{(1)}\right)^2}\right]=(1-m_\star)\frac{\mathsf{E}U^2V^{m_\star-2}}{\mathsf{E}V^{m_\star}},$$

where $U=\sinh(\beta g), V=\cosh(\beta g)$ for g standard gaussian (remark that $U^2V^{m_\star-2}=\tanh^2(\beta g)\cosh(\beta g)^{m_\star}$), and therefore

$$\lim_{N\to\infty}(II)=-2(1-m_\star)q_\star^2.\tag{2.41}$$

Finally, for the term (I) we proceed analogously, with $U=\sinh(\beta g_1)\sinh(\beta g_2)$, $V=\cosh(\beta g_1)\cosh(\beta g_2)$ (and g_1,g_2 standard, independent gaussians):

$$\begin{aligned}\lim_{N\to\infty}(I)&=(1-m_\star)\frac{\mathsf{E}[U^2V^{m_\star-2}]}{\mathsf{E}[V_{m_\star}]}\\&=(1-m_\star)\frac{\mathsf{E}\left[\tanh^2(\beta g_1)\tanh^2(\beta g_2)\cosh(\beta g_1)^{m_\star}\cosh(\beta g_2)^{m_\star}\right]}{\mathsf{E}\left[\cosh(\beta g_1)^{m_\star}\cosh(\beta g_2)^{m_\star}\right]}\\&=(1-m_\star)q_\star^2,\end{aligned}\tag{2.42}$$

the last step by the independence of g_1 and g_2. Combining the information gathered in (2.40), (2.41) and (2.42) yields claim $i)$ of the Proposition. Claim $ii)$ involves similar considerations. □

PROOF OF PROPOSITION 2.8. Consider for the moment the β-system only. Denote by $f(\beta)$ the free energy and by $G_{m_\star(\beta)}$ the associated extremal measure solving the GVP. For $\varepsilon > 0$, set $B_{\beta,\varepsilon} \stackrel{\text{def}}{=} B_\varepsilon(G_{m_\star(\beta)}) \subset \mathcal{M}_1^+(\mathbb{R}^2)$ for the open ball of radius ε and center $G_{m_\star(\beta)}$. For $\alpha \in \Sigma_N$ we denote by $L_{N,\alpha}$ the empirical measures associated to the free energies of the pure states.

We first claim that given $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\mathbb{P}\left[\mathcal{G}_{\beta,N}\left(\alpha \in \Sigma_N : L_{N,\alpha} \notin B_{\beta,\varepsilon}\right) \geq e^{-\delta N}\right] \leq e^{-\delta N}. \tag{2.43}$$

To see this, first observe that uniqueness of the maximizers solving the GVP implies that, with $\phi(x_1, x_2) = \beta x_1 + \log\cosh(\beta x_2)$ and μ standard bivariate gaussian,

$$f(\beta,\varepsilon) \stackrel{\text{def}}{=} \sup\left\{\mathsf{E}_\nu[\phi] - H(\nu \mid \mu) : H(\nu \mid \mu) \leq \log 2, \nu \notin B_{\beta,\varepsilon}\right\} < f(\beta). \tag{2.44}$$

Using the same argument as in the proof of Theorem 2.1 we get

$$\lim_{N\to\infty} \frac{1}{N} \log 2^{-N} \sum_{\alpha : L_{N,\alpha} \notin B_{\beta,\varepsilon}} e^{\beta X_\alpha + \sum_{i=1}^N \log\cosh(\beta g_{\alpha,i})} \leq f(\beta,\varepsilon),\ \mathbb{P} - a.s.$$

Using the variance estimate from (2.13) and the Tchebychev inequality, it is easily seen that

$$\mathbb{P}[\Lambda_N^c(\beta,\varepsilon)] \leq \exp(-\delta N), \tag{2.45}$$

for some $\delta > 0$, where $\Lambda_N(\beta,\varepsilon)$ is the event

$$\begin{aligned}
\Big\{2^{-N} \sum_{\alpha : L_{N,\alpha} \notin B_{\beta,\varepsilon}} e^{\beta X_\alpha + \sum \log\cosh(\beta g_{\alpha,i})} \leq f(\beta,\varepsilon) + \frac{\eta}{3},& \\
2^{-N} \sum_{\alpha} e^{\beta X_\alpha + \sum \log\cosh(\beta g_{\alpha,i})} \geq f(\beta,\varepsilon) + \frac{2}{3}\eta\Big\},& \qquad (2.46)\\
\eta \stackrel{\text{def}}{=} f(\beta) - f(\beta,\varepsilon),&
\end{aligned}$$

which clearly implies (2.43).

A similar statement evidently holds for the β'-system.

We now observe that to given $\varepsilon, \delta > 0$,

$$\begin{aligned}
&\mathbb{P}\left[\mathcal{G}_{\beta,N} \otimes \mathcal{G}_{\beta',N}\left(\delta_{\alpha,\alpha'} = 1\right) \geq e^{-\delta N}\right] \leq \\
&\quad \leq \mathbb{P}\left[\mathcal{G}_{\beta,N}(\alpha : L_{N,\alpha} \notin B_{\beta,\varepsilon}) \geq e^{-\delta N}\right] + \mathbb{P}\left[\mathcal{G}_{\beta',N}(\alpha' : L_{N,\alpha'} \notin B_{\beta',\varepsilon}) \geq e^{-\delta N}\right] + \\
&\quad\quad + \mathbb{P}\left[\mathcal{G}_{\beta,N} \otimes \mathcal{G}_{\beta',N}\left(\alpha : L_{N,\alpha} \in B_{\beta,\varepsilon} \cap B_{\beta',\varepsilon}\right) \geq e^{-\delta N}\right].
\end{aligned} \tag{2.47}$$

For $\beta \neq \beta'$ we clearly have $G_{m_\star(\beta)} \neq G_{m_\star(\beta')}$. We may therefore choose ε small enough such that $B_{\beta,\varepsilon} \cap B_{\beta',\varepsilon} = \emptyset$ and the last term on the r.h.s. above drops out. By (2.43), the first and second term are exponentially small (in N). This proves part a) of the Proposition 2.8.

As for claim b), we observe that

$$\mathbb{E}\left\langle \delta_{\alpha\neq\alpha'} q(\sigma,\sigma')^2 \right\rangle^{\otimes 2}_{\beta,\beta',N} = \mathbb{E}\left\langle q(\sigma,\sigma')^2 \right\rangle^{\otimes 2}_{\beta,\beta',N} - \mathbb{E}\left\langle \delta_{\alpha=\alpha'} q(\sigma,\sigma')^2 \right\rangle^{\otimes 2}_{\beta,\beta',N}. \tag{2.48}$$

Since $q(\sigma,\sigma')^2 \leq 1$ for all σ,σ', by claim a) of this Proposition the second term on the r.h.s in (2.48) is in the limit $N \to \infty$ vanishing. As for the first term on the r.h.s of (2.48), by symmetry and obvious bounds we have

$$\mathbb{E}\left\langle q(\sigma,\sigma')^2 \right\rangle^{\otimes 2}_{\beta,\beta',N} = \{1 + O(1/N)\}\, \mathbb{E}\left\langle \sigma_1\sigma_1'\sigma_2\sigma_2' \right\rangle^{\otimes 2}_{\beta,\beta',N} + O(1/N). \tag{2.49}$$

Let us now set $w_\alpha^{(1,2,\beta)} \stackrel{\text{def}}{=} \exp\left[\beta X_\alpha + \sum_{i=3}^N \log\cosh(\beta g_{\alpha,i}) - a_N(\beta)\right]$ for the Boltzmann weight of the pure state α with a cavity in the sites $i = 1,2$ associated to the β-system, and $a_N(\beta)$ being the centering constant from (2.28) specialized to the setting. Analogously, we write $w_\alpha^{(1,2,\beta')} \stackrel{\text{def}}{=} \exp\left[\beta' X_\alpha + \sum_{i=3}^N \log\cosh(\beta' g_{\alpha,i}) - a_N(\beta')\right]$ in case of the β'-system. With this notations in mind, we write the expectation on the r.h.s of (2.48) as

$$\begin{aligned} &\mathbb{E}\left\langle \sigma_1\sigma_1'\sigma_2\sigma_2' \right\rangle^{\otimes 2}_{\beta,\beta',N} = \\ &= \mathbb{E}\left[\frac{\sum_\alpha \sinh(\beta g_{\alpha,1})\sinh(\beta g_{\alpha,2}) w_\alpha^{(1,2,\beta)}}{\sum_\alpha \cosh(\beta g_{\alpha,1})\cosh(\beta g_{\alpha,2}) w_\alpha^{(1,2,\beta)}} \times \frac{\sum_{\alpha'} \sinh(\beta' g_{\alpha',1})\sinh(\beta' g_{\alpha',2}) w_{\alpha'}^{(1,2,\beta')}}{\sum_{\alpha'} \cosh(\beta' g_{\alpha',1})\cosh(\beta' g_{\alpha',2}) w_{\alpha'}^{(1,2,\beta')}} \right]. \end{aligned} \tag{2.50}$$

By Proposition 2.11.b) the Point Process associated to the collection $(w_\alpha^{(1,2,\beta)})$, that of the β-system, converges weakly to a $PPP(t \mapsto t^{-m_\star(\beta)-1})$, while the Point Process associated to the β'-system converges to a $PPP(t \mapsto t^{-m_\star(\beta')-1})$. On the other hand, using similar arguments as in the proof of claim a) it is not difficult to see that the limiting point processes are in fact independent. [Given a compact set $K \subset \mathbb{R}_+$, the $\mathbb{P}$-probability to find a configuration $\alpha \in \Sigma_N$ such that $w_\alpha^{(1,2,\beta)} \in K$ and simultaneously $w_\alpha^{(1,2,\beta')} \in K$ is exponentially small in N.] Hence, the r.h.s of (2.50) converges with $N \to \infty$ to the product

$$\mathsf{E}\left[\frac{\sum_\alpha \sinh(\beta g_{\alpha,1})\sinh(\beta g_{\alpha,2}) w_\alpha}{\sum_\alpha \cosh(\beta g_{\alpha,1})\cosh(\beta g_{\alpha,2}) w_\alpha} \right] \times \mathsf{E}\left[\frac{\sum_\alpha \sinh(\beta' g_{\alpha,1})\sinh(\beta' g_{\alpha,2}) w_\alpha'}{\sum_\alpha \cosh(\beta' g_{\alpha,1})\cosh(\beta' g_{\alpha,2}) w_\alpha'} \right], \tag{2.51}$$

with (w_α) a PPP of density $t^{-m_\star(\beta)-1}dt$, and (w_α') a PPP of density $t^{-m_\star(\beta')-1}dt$ respectively. By the first formula in Lemma 2.14, both expectations are easily seen to be equal to zero. This settles claim b) of Proposition 2.8. □

3. On a cavity field Perturbation of the GREM

3.1. Introduction and outline

We extend the abstract approach introduced in Chapter 2 to mean field models of spin glasses displaying a (not necessarily linear) GREM-structure. This allows in particular for a straightforward computation of the free energy associated to a GREM under the influence of an extensive cavity field. Our interest in the latter stems from the approach to the Sherrington-Kirkpatrick model under the cavity dynamics as is perhaps best seen in the elegant framework of Aizenman-Sims-Starr [**1**].

The new and crucial observation in the whole treatment concerns the rôle played by certain empirical measures associated to the energy levels, to which one can apply succesfully the socalled Second Moment Method with truncation. Working with empirical measures naturally leads to infinite dimensional problems, an overhall delicate infrastructure, and related fiddling issues. The outcome is however well worth the effort. First of all, the usual assumptions on the random medium, i.e. the typical gaussian character of the interactions, are by no means necessary, the methods working just the same for any reasonable underlying distribution. Secondly, the infinite dimensional problems obtained in the large N-limit turn out to be quite easy to handle, and reducible to finite dimensional variational problems; these can then be reformulated in terms of the Parisi unorthodox minimization. Moreover, and most importantly, the dimensional reduction goes through the identification of an order parameter which coincides with that of the Parisi Theory, the latter acquiring a new, appealing interpretation.

3.2. Mean field models of GREM-type

We consider a measurable space $(S, \mathcal{S})$, write $\mathcal{M}_1^+(S, \mathcal{S})$, or $\mathcal{M}_1^+(S)$ for short, for the set of probability measures on $(S, \mathcal{S})$. If $\mu, \nu \in \mathcal{M}_1^+(S)$, then

$$H(\nu \mid \mu) \stackrel{\text{def}}{=} \begin{cases} \int \log \frac{d\nu}{d\mu} d\nu, & \text{if } \nu \ll \mu \\ \infty & \text{otherwise,} \end{cases}$$

is the usual relative entropy.

For $n \in \mathbb{N}$, $\alpha = (\alpha_1, \ldots, \alpha_n)$ with $1 \leq \alpha_i \leq 2^{\gamma_i N}$, $\sum_i \gamma_i = 1$, and $1 \leq i \leq N$, let

$$X_{\alpha,i} = \left(X^1_{\alpha_1,i}, X^2_{\alpha_1,\alpha_2,i}, \ldots, X^K_{\alpha_1,\alpha_2,\ldots,\alpha_K,i}\right)$$

where the X^j are indepedent, taking values in some Polish Space $(S, \mathcal{S})$ with distribution μ_j. Set

$$L_{N,\alpha} = \frac{1}{N} \sum_{i=1}^{N} \delta_{\left(X^1_{\alpha_1,i}, X^2_{\alpha_1,\alpha_2,i}, \ldots, X^n_{\alpha_1,\alpha_2,\ldots,\alpha_n,i}\right)},$$

which is a random element in $\mathcal{M}_1^+(S^n)$.

Let $\Phi : \mathcal{M}_1^+(S^n) \to \mathbb{R}$ be continous and set $Z_N \stackrel{\text{def}}{=} 2^{-N} \sum_\alpha \exp\left[N\Phi(L_{N,\alpha})\right]$.

Theorem 3.1. *It holds:*

$$\begin{aligned}\lim_{N\to\infty} \frac{1}{N}\log Z_N = \\ = \sup\left\{\Phi(\nu) - H(\nu \mid \mu) :\ H\left(\nu^{(j)} \mid \mu^{(j)}\right) \leq \Gamma_j \log 2,\ j = 1, \dots, n\right\},\end{aligned} \tag{3.1}$$

$\mathbb{P}-a.s.$, where $\nu^{(j)}$ is the marginal measure of $\nu \in \mathcal{M}_1^+(S^n)$ on the first j components, $\mu^{(j)} = \mu_1 \otimes \cdots \otimes \mu_j, \mu \stackrel{def}{=} \mu^{(n)}$, and $\Gamma_j \stackrel{def}{=} \sum_{r=1}^j \gamma_j$.

We now specialize to linear functionals of the form $\Phi(\nu) = \int \phi(x)\nu(dx)$, for $\nu \in \mathcal{M}_1^+(S^n)$. We also write $\mathsf{E}_\nu[\phi] \stackrel{\text{def}}{=} \int \phi(x)\nu(dx)$, and

$$\mathsf{Gibbs}(\phi, \nu, \mu) \stackrel{\text{def}}{=} \mathsf{E}_\nu[\phi] - H(\nu \mid \mu).$$

Remark that if ϕ is such that the assumptions in Theorem 3.1 are satisfied, the variational problem on the r.h.s of (3.1)

$$= \sup\left\{\mathsf{Gibbs}(\phi, \nu, \mu) : H\left(\nu^{(j)} \mid \mu^{(j)}\right) \leq \Gamma_j \log 2,\ j = 1, \dots, n\right\},$$

an expression we will refer to as Gibbs Variational Principle (in full analogy with the finite dimensional case), GVP for short. The analogy also motivates the following definition of *Generalized Gibbs measure* below, for which we need some additional notation: for $\boldsymbol{m} = (m_1, m_2, \dots, m_n) \in \mathbb{R}^n$, ϕ and μ as above, we define the following sequence of functions by recursion

$$\phi_n \stackrel{\text{def}}{=} \phi, \quad \phi_{i-1}(x_1, \dots, x_{i-1}) \stackrel{\text{def}}{=} \frac{1}{m_i} \log \int \phi_i^{m_i}(x_1, \dots, x_{i-1}, y)\mu_i(dy)$$

for $i = n, \dots, 1$.

Definition 3.2. *A generalized Gibbs measure $G = G(\phi, \boldsymbol{m}, \mu)$ on $\mathbb{R}^n$ is of the form*

$$\frac{dG}{d\mu}(x_1, \dots, x_n) = \prod_{j=1}^n G_j(x_1, \dots, x_j), \qquad G_j(x_1, \dots, x_j) = \frac{\exp m_j\phi_j}{\mathsf{E}_{\mu_j}(\exp m_j\phi_j)},$$

We also set $\mathcal{P}(\phi, \nu, m_1, \dots, m_n) \stackrel{\text{def}}{=} \phi_0$, and introduce the following functional

$$\mathsf{Parisi}(\phi, \boldsymbol{m}, \mu) \stackrel{\text{def}}{=} \sum_{i=1}^n \frac{\gamma_i \log 2}{m_i} + \mathcal{P}\left(\phi, \mu, m_1, \dots, m_n\right) - \log 2. \tag{3.2}$$

As for the reason of the latter terminology, we point out that an abstract version of the RSB-computations à la Parisi (under the additional assumption of self-averaging) would lead to an "abstract free energy" given by

$$\lim_{N\to\infty} \frac{1}{N}\log Z_N = \min_{\boldsymbol{m}\in\triangle} \mathsf{Parisi}(\phi, \boldsymbol{m}, \mu)\ \mathbb{P}-a.s.,$$

where $\triangle \stackrel{\text{def}}{=} \{\boldsymbol{m} \in \mathbb{R}^n : 0 \leq m_1 \leq m_2 \leq \cdots \leq m_n \leq 1\}$. Although the RSB-scheme is highly non rigorous, the minimization on the r.h.s of (3.2) makes perfect sense, and it comes perhaps as a surprise that it gives also the correct answer:

Theorem 3.3. *Within the setting of Theorem 3.1,*

i) *there exists a unique extremal measure which solves the GVP;*
ii) *it is a generalized Gibbs measure* $G(\phi, \underline{\boldsymbol{m}}, \mu)$ *with "order parameter"*

$$\underline{\boldsymbol{m}} = \arg \min_{\boldsymbol{m} \in \triangle} \mathsf{Parisi}(\phi, \boldsymbol{m}, \mu), \tag{3.3}$$

iii) *With this choice of order parameter,* $\mathsf{Gibbs}(\phi, G(\phi, \underline{\boldsymbol{m}}), \mu) = \mathsf{Parisi}(\phi, \underline{\boldsymbol{m}}, \mu)$.

Theorem 3.3 not only provides a link between the Parisi minimization and the GVP, but it also provides an novel interpretation of the order parameter of the Parisi Theory, the vector $\underline{\boldsymbol{m}}$, in terms of the sequence of (inverse of) temperatures associated to the extremal Generalized Gibbs measure solving the GVP.

3.3. An application, the GREM+Cavity

For $N \in \mathbb{N}$ we define $\Sigma_N^{\text{grem}} \stackrel{\text{def}}{=} \{1, \dots, 2^N\}$ and for $i \in I = \{1, \dots, n\}, \Sigma_{N,i}^{\text{grem}} = \Sigma_{N\gamma_i}^{\text{grem}}$, where for notational convenience we assume that $N\gamma_i$ is an integer. The spin configurations of the GREM are

$$\alpha = (\alpha_1, \dots, \alpha_n), \ \alpha_i \in \Sigma_{N,i}^{\text{grem}},$$

i.e. we identify $\Sigma_N^{\text{grem}} = \Sigma_{N,1}^{\text{grem}} \times \cdots \times \Sigma_{N,n}^{\text{grem}}$. The Hamiltonian of the GREM is then given by

$$X_\alpha \stackrel{\text{def}}{=} X_{\alpha_1} + X_{\alpha_1,\alpha_2} + \dots X_{\alpha_1,\dots,\alpha_n},$$

where the $X_{\alpha_1,\dots,\alpha_i}, i \in I$ are independent, centered gaussians of variance $a_i N$. Clearly, the X_α are no longer indepedent,

$$\mathbb{E}(X_\alpha X_{\alpha'}) = N \left\{ \sum_{i=1}^{q(\alpha,\alpha')} a_i \right\}, \ q(\alpha, \alpha') \stackrel{\text{def}}{=} \max\{i \in I : (\alpha_1, \dots, \alpha_i) = (\alpha_1', \dots, \alpha_i')\}$$

We also take a second family of centered gaussian random variables

$$\left(g_{\alpha_1,\alpha_2,\dots,\alpha_i,j}; \ \alpha_i \in \Sigma_{N,i}^{\text{grem}}, 1 \le i \le n, 1 \le j \le N \right),$$

independent of the (X_α) and covariance given by

$$\mathbb{E}\left[g_{\alpha_1,\dots,\alpha_i,j} g_{\alpha_1',\dots,\alpha_i',j'} \right] = \begin{cases} a_i & \text{if } j = j', \alpha_1, \dots, \alpha_i = \alpha_1', \dots, \alpha_i' \\ 0 & \text{otherwise,} \end{cases}$$

and write $g_{\alpha,j} = g_{\alpha_1,j} + g_{\alpha_1,\alpha_2,j} + \cdots + g_{\alpha_1,\dots,\alpha_n,j}$.

For Ising spin configurations $\sigma = (\sigma_1, \dots, \sigma_N) \in \{\pm 1\}^N$ we set the hamiltonian of the "GREM+Cavity" to be

$$X_{\alpha,\sigma} = X_\alpha + \sum_{i=1}^{N} g_{\alpha,i}\sigma_i, \quad \text{for } (\alpha, \sigma) \in \Sigma_N^{\text{grem}} \times \{\pm 1\}^N. \tag{3.4}$$

Remark that

$$\mathbb{E}[X_{\alpha,\sigma}X_{\alpha',\sigma'}] = N \sum_{i=1}^{q(\alpha,\alpha')} \Big\{a_i + q_N(\sigma,\sigma')a_i\Big\}, \tag{3.5}$$

where $q_N(\sigma,\sigma') = 1/N\sum_{i=1}^N \sigma_i\sigma'_i$ is the usual "overlap".

We write $\mathrm{tr}(\cdot)$ for averaging over $\Sigma_N^{\mathrm{grem}} \times \{\pm 1\}^N$, i.e. the coin-tossing expectation if we identify $\Sigma_N^{\mathrm{grem}} \times \{\pm 1\}^N$ with $\{H,T\}^{2N}$) and for a function $x : \Sigma_N^{\mathrm{grem}} \times \{\pm 1\}^{N/2} \to \mathbb{R}$, set

$$Z_N(\beta,x) \stackrel{\mathrm{def}}{=} \mathrm{tr}\exp[\beta x], \quad F_N(\beta,x) = \frac{1}{N}\log Z_N(\beta,x),$$

and define the usual finite N partition function and free energy

$$Z_N(\beta) \stackrel{\mathrm{def}}{=} Z_N(\beta,X), \quad F_N(\beta) \stackrel{\mathrm{def}}{=} F_N(\beta,X), \quad f_N(\beta) \stackrel{\mathrm{def}}{=} \mathbb{E}\left[F_N(\beta,X)\right],$$

where the hamiltonian (3.4) is interpreted as a random function $X : \Sigma_N^{\mathrm{grem}} \times \{\pm 1\}^N \to \mathbb{R}$.

Theorem 3.4. i) *The limiting free energy $f(\beta) \stackrel{def}{=} \lim_{N\to\infty} f_N(\beta)$ exists, and is the almost sure limit of $F_N(\beta)$.*

ii) *Let μ_i be standard bivariate gaussian distributions on $\mathcal{M}\left(\mathbb{R}^2\right), i = 1,\dots,n$ and set $\mu \stackrel{def}{=} \otimes_{i=1}^n \mu_i$. Set also*

$$\phi\Big((x_1,y_1),\dots,(x_n,y_n)\Big) = \left\{\sum_{i=1}^n \beta\sqrt{a_i}x_i\right\} + \log\cosh(\beta\sqrt{a_1}y_1 + \cdots + \beta\sqrt{a_n}y_n).$$

Then, the following representation holds:

$$f(\beta) = \sup_{\nu\in\mathcal{M}(\mathbb{R}^{2n})} \left\{\mathsf{Gibbs}(\phi,\nu,\mu) : H\left(\nu^{(j)} \mid \mu^{(j)}\right) \leq \Gamma_j \log 2,\ j \leq n\right\} \tag{3.6}$$

We are not able to describe properly the behavior of the order parameter $\underline{\boldsymbol{m}}$ as a function of β and $(a_i,\gamma_i, i\in I)$. Here we summarize what we can say about the phase diagram of the GREM+Cavity: Having fixed the underlying parameters (a_i,γ_i), it is easily seen that for small enough β, the minimum in Parisi minimization is achieved in $\underline{\boldsymbol{m}} = (1,1,\dots,1)$, in which case the limiting free energy reads $f(\beta) = \beta^2$ and thus equals the annealed free energy $(1/N)\log\mathbb{E}[Z_N(\beta,X)]$. This is the socalled replica symmetric regime, RS for short. On the other hand, by the same considerations as in the REM+Cavity, one can easily prove that for large enough β replica symmetry is broken: there exists $\tilde{m} \in (0,1)$ such that the Parisi functional evaluted at $m_1 = m_2 = \cdots = m_n = \tilde{m}$ is strictly less than its value in $m_1 = m_2 = \cdots = m_n = 1$, and thus $f(\beta) < \beta^2$. This also implies that the following holds:

K-RSB. There exists $K \in \{1,\dots,n\}$, and a strictly increasing sequence of reals $0 < m^{(1)} < m^{(2)} < \cdots < m^{(K)} < m^{(K+1)} = 1$, integers $0 = j_0 < j_1 < j_2 < \cdots < j_K < j_{K+1} = n$ such that the choice $\underline{m_l} \equiv m^{(r)}$ for $l = j_{r-1}+1,\dots,j_r$ and $r = 1,\dots,K+1$ is optimal for the Parisi minimization, and thus for the GVP as well.

Remark 3.5. *The limiting free energy of the GREM+Cavity assumes in case of K-RSB a form reminiscent (but not identical) of the SK-solution. In fact, by simple computations one can see that, letting* $a^{(r)} = \sum_{i=j_{r-1}+1}^{j_r} a_i$, $\gamma^{(r)} \stackrel{def}{=} \sum_{i=j_{r-1}}^{j_r} \gamma_i$, *and*

$$\cosh(\beta; x_1, \ldots, x_k) \stackrel{def}{=} \cosh\left(\beta\sqrt{a^{(1)}}x_1 + \cdots + \beta\sqrt{a^{(K)}}x_K\right),$$

then, for $\mu = \otimes_{i=1}^n \mu_i$ *with* μ_i *standard gaussians on* $\mathbb{R}$,

$$\begin{aligned} f(\beta) = &\sum_{r=1}^{K}\left[\frac{\log 2}{m^{(r)}}\gamma^{(r)} + \frac{\beta^2}{2}a^{(r)}m^{(r)}\right] + \beta^2 a^{(K+1)} \\ &+ \frac{1}{m^{(1)}}\log \mathsf{E}_{\mu_1}\left[\ldots\left[\mathsf{E}_{\mu_K}\left[\cosh\left(\beta;(\cdot)\right)^{m^{(K)}}\right]^{m^{(K-1)}/m^{(K)}}\right]\ldots\right] - \gamma^{(n)}\log 2. \end{aligned} \tag{3.7}$$

Remark also that in the terminology usually adopted in the case of the SK-model, the "order parameter" would rather be the non decreasing function

$$\begin{aligned} x : [0,1] &\to [0,1], \\ q &\mapsto x(q) = m^{(r)}, \ \textit{for } q \in [q_{r-1}, q_r). \end{aligned} \tag{3.8}$$

and $q_r \stackrel{def}{=} \sum_{j=1}^r a^{(j)}$.

3.4. Proofs of the main results

We begin with some general considerations pertaining to the relative entropies.

If $A \in \mathcal{S}$, we put $H(A \mid \mu) \stackrel{\text{def}}{=} \inf_{\nu\in A} H(\nu \mid \mu)$. If S is a Polish Space, and $\mathcal{S}$ its Borel σ-field, then it is well known that $\nu \to H(\nu \mid \mu)$ is lower semicontinuous in the weak topology, i.e. if $\nu_n \to \nu$ weakly, then

$$H(\nu \mid \mu) \le \liminf_{n\to\infty} H(\nu_n \mid \mu).$$

This follows from the representation

$$H(\nu \mid \mu) = \sup_{u\in\mathcal{U}}\left[\int u\,d\nu - \log\int e^u\,d\mu\right], \tag{3.9}$$

where $\mathcal{U}$ is the set of bounded continuous functions $S \to \mathbb{R}$.
For $(S, \mathcal{S}), (S', \mathcal{S}')$ two Polish Spaces, and $\nu \in \mathcal{M}_1^+(S \times S')$. If $\mu \in \mathcal{M}_1^+(S)$, $\mu' \in \mathcal{M}_1^+(S')$ we have,

$$H\left(\nu \mid \mu\otimes\mu'\right) = H\left(\nu^{(1)} \mid \mu\otimes\mu'\right) + H\left(\nu \mid \nu^{(1)}\otimes\mu'\right),$$

where $\nu^{(1)}$ is the first marginal of ν on S.

Lemma 3.6. $H(\nu \mid \nu^{(1)} \otimes \mu')$ *is a lower semicontinuous function of* ν *in the weak topology.*

Proof. Applying (3.9) to

$$H(\nu \mid \nu^{(1)} \otimes \mu') = \sup_{u \in \mathcal{U}} \left[\int u d\nu - \log \int e^u d\left(\nu^{(1)} \otimes \mu'\right) \right],$$

where $\mathcal{U}$ denotes the set of bounded continuous functions $S \times S' \to \mathbb{R}$. For any fixed $u \in \mathcal{U}$, both functions $\nu \to \int u\, d\nu$ and $\nu \to \log \int e^u d\left(\nu^{(1)} \otimes \mu'\right)$ are continuous, and from this the desired semicontinuity property follows. □

We will need the following "relative" version of Sanov's theorem. Consider three independent sequences of i.i.d. random variables $(X_i), (Y_i), (Z_i)$, taking values in three Polish spaces S, S', S'' and with laws μ, μ', μ''. We consider the empirical processes

$$L_N \stackrel{\text{def}}{=} \frac{1}{N} \sum_{i=1}^{N} \delta_{(X_i, Y_i)}, \quad R_N = \frac{1}{N} \sum_{i=1}^{N} \delta_{X_i, Z_i}.$$

The pair (L_N, R_N) takes values in $\mathcal{M}_1^+(S \times S') \times \mathcal{M}_1^+(S \times S'')$.

Lemma 3.7. *The sequence (L_N, R_N) satisfies a full LDP with good rate function*

$$J(\nu, \theta) = \begin{cases} H\left(\nu^{(1)} \mid \mu\right) + H\left(\nu \mid \nu^{(1)} \otimes \mu'\right) + H\left(\theta \mid \theta^{(1)} \otimes \mu''\right), & \text{if } \nu^{(1)} = \theta^{(1)} \\ \infty & \text{otherwise.} \end{cases}$$

Proof. We apply the standard Sanov theorem to the empirical measure

$$M_N = \frac{1}{N} \sum_{i=1}^{N} \delta_{(X_i, Y_i, Z_i)} \in \mathcal{M}_1^+(S \times S' \times S'').$$

We use the two natural projections $p : S \times S' \times S'' \to S \times S'$ and $q : S \times S' \times S'' \to S \times S''$. Then $(L_N, R_N) = M_N(p, q)^{-1}$, and by continuous projection, we get that (L_N, R_N) satisfies a good LDP with rate function

$$J'(\nu, \theta) = \inf \left\{ H(\rho \mid \mu \otimes \mu' \otimes \mu'') : \ \rho p^{-1} = \nu, \rho q^{-1} = \theta \right\}$$

It only remains to identify this rate function with the function J given above.

Clearly $J'(\nu, \theta) = \infty$ if $\nu^{(1)} \neq \theta^{(1)}$. Therefore, assume $\nu^{(1)} = \theta^{(1)}$. If we choose ρ to have marginal $\nu^{(1)} = \theta^{(1)}$ on S, and the conditional distribution on S', S'' given the first projection is the product of the conditional distirbutions of ν and θ, then evidently

$$H(\rho \mid \mu \otimes \mu' \otimes \mu'') = H\left(\nu^{(1)} \mid \mu\right) + H\left(\nu \mid \nu^{(1)} \otimes \mu'\right) + H\left(\theta \mid \theta^{(1)} \otimes \mu''\right),$$

and therefore $J \geq J'$. On the other hand, Jensen's inequality shows

$$J(\nu, \theta) \leq H\left(\rho \mid \mu \otimes \mu' \otimes \mu''\right),$$

for any ρ satisfying $\rho p^{-1} = \nu, \rho q^{-1} = \theta$. □

We now step back to the setting of Theorem 3.1: we are given independent sequences $(X^j_{\alpha_1, \dots, \alpha_j, i})$ of independent random variables with distribution μ_j for $j = 1, \dots, n$. We emphasize that henceforth $\mu = \mu_1 \otimes \cdots \otimes \mu_n$ and $\mu^{(k)}$ will denote the marginal on the first k-components. Moreovoer, for $\alpha = (\alpha_1, \ldots, \alpha_n)$, we write $\alpha^{(j)} = (\alpha_1, \ldots, \alpha_j)$ and set

$$L^{(j)}_{N,\alpha^{(j)}} = \frac{1}{N} \sum_{i=1}^{N} \delta_{\left(X^1_{\alpha_1,i}, X^2_{\alpha_1,\alpha_2,i}, \ldots, X^j_{\alpha_1,\ldots,\alpha_j,i}\right)},$$

for $j \leq n$, which is the marginal of $L_{N,\alpha}$ on S^j. With the notation

$$X^{(j)}_{\alpha^{(j)},i} \stackrel{\text{def}}{=} \left(X^1_{\alpha_1,i}, \ldots, X^j_{\alpha_1,\ldots,\alpha_j,i}\right)$$
$$\hat{X}^{(j)}_{\alpha^{(j)},i} \stackrel{\text{def}}{=} \left(X^{j+1}_{\alpha_1,\ldots,\alpha_{j+1},i}, \ldots, X^n_{\alpha_1,\ldots,\alpha_n,i}\right)$$

we can write

$$L_{N,\alpha} \stackrel{\text{def}}{=} \frac{1}{N} \sum_{i=1}^{N} \delta_{\left(X^{(j)}_{\alpha^{(j)},i}, \hat{X}^{(j)}_{\alpha^{(j)},i}\right)} \tag{3.10}$$

For $A \subset \mathcal{M}^+_1(S^n)$ we put $M_N(A) \stackrel{\text{def}}{=} \sharp\{\alpha : L_{N,\alpha} \in A\}$.

Lemma 3.8. *Assume $\nu \in \mathcal{M}^+_1(S^n)$ satisfies $H(\nu \mid \mu) < \infty$, and let V be an open neighborhood of ν, and $\varepsilon > 0$. Then there exists an open neighborhood U of ν, $U \subset V$, and $\delta > 0$ such that*

$$\mathbb{P}\Big[M_N(U) \geq \exp\left[N(\log 2 - H(\nu \mid \mu) + \varepsilon)\right]\Big] \leq e^{-\delta N},$$

if N is large enough.

Proof. We can choose a sequence $r_k > 0, r_k \downarrow 0$ with $H(B_{r_k}(\nu \mid \mu)) = H(\mathrm{cl}B_{r_k}(\nu) \mid \mu) \to H(\nu \mid \mu)$, where $B_r(\nu)$ denotes the open r-ball around ν in one of the standard metrics, e.g. the Prohorov metric. By Sanov's theorem we get for k large enough, such that $B_{r_k}(\nu) \subset V$ and $H(B_{r_k}(\nu) \mid \mu) \geq H(\nu \mid \mu) - \varepsilon/4$,

$$\mathbb{P}\Big[L_{N,\alpha} \in B_{r_k}(\nu)\Big] \leq \exp\left[N(-H(\nu \mid \mu) + \varepsilon/2)\right],$$

for N large enough, and therefore

$$\mathbb{E}\Big[M_N\big(B_{r_k}(\nu)\big)\Big] \leq \exp\left[N(\log 2 - H(\nu \mid \mu) - \varepsilon/2)\right],$$

from which the statement follows by Markov inequality (with $\delta = \varepsilon/2$). $\square$

Lemma 3.9. *Assume $\nu \in \mathcal{M}^+_1(S^n)$ satisfies $H\left(\nu^{(j)} \mid \mu^{(j)}\right) > \Gamma_j \log 2$ for some $j \leq n$, and let V be an open neighborhood of ν. Then there is an open neighborhood U of ν, $U \subset V$ and $\delta > 0$ such that*

$$\mathbb{P}\Big[M_N(U) \neq 0\Big] \leq e^{-\delta N}$$

for large enough N.

Proof. We first choose a neighborhood U' of $\nu^{(j)}$ in S^j such that $H(\mathrm{cl}\,U' \mid \mu^{(j)}) = H(U' \mid \mu^{(j)}) > \Gamma_j \log 2$, say $H(\mathrm{cl}\,U' \mid \mu^{(j)}) \geq \Gamma_j \log 2 + \eta$. Then we put

$$U \stackrel{\text{def}}{=} \left\{\nu \in \mathcal{M}^+_1(S^n) : \nu \in V, \nu^{(j)} \in U'\right\}.$$

If $L_{N,\alpha} \in U$ then $L^{(j)}_{N,\alpha^{(j)}} \in U'$,

$$\begin{aligned}\mathbb{P}\left[\exists\alpha:\ L_{N,\alpha}\in U\right] &\le \mathbb{P}\left[\exists\alpha^{(j)}:\ L^{(j)}_{N,\alpha^{(j)}}\in U'\right]\\ &\le 2^{\Gamma_j N}\mathbb{P}\left[L^{(j)}_{N,\alpha^{(j)}}\in U'\right] \le 2^{\gamma_j N}\exp\left[-NH\left(\mathrm{cl}\,U' \mid \mu^{(j)}\right)+N\eta/2\right]\\ &\le 2^{\Gamma_j N}\exp\left[-N\Gamma_j\log 2 - N\eta/2\right] = e^{-N\eta/2}.\end{aligned}$$

This proves the claim. □

Lemma 3.10. *Assume that $\nu \in \mathcal{M}_1^+(S^n)$ satisfies $H\left(\nu^{(j)} \mid \mu^{(j)}\right) < \Gamma_j \log 2$ for all j, and let V be an open neighborhood of ν, and $\varepsilon > 0$. Then there exists an open neighborhood U of ν, $U \subset V$, and a $\delta > 0$ such that*

$$\mathbb{P}\Big[M_N(U) \le \exp\left[N\left(\log 2 - H(\nu \mid \mu) - \varepsilon\right)\right]\Big] \le e^{-\delta N}.$$

Proof. We claim that we can find the U such that

$$\mathbb{E}\left[M_N(U)\right] \ge \exp\left[N\left(\log 2 - H(\nu \mid \mu) - \varepsilon\right)\right], \tag{3.11}$$

and for some $\delta > 0$

$$\operatorname{var}\left[M_N(U)\right] \le e^{-2N\delta}\left\{\mathbb{E}\left[M_N(U)\right]\right\}^2 \tag{3.12}$$

for large N. Then,

$$\begin{aligned}&\mathbb{P}\Big[M_N(U) \le \exp\left[N(\log 2 - H(\nu \mid \mu) - \varepsilon)\right]\Big]\\ &\qquad\qquad \le \mathbb{P}\left[\left|M_N(U) - \mathbb{E}M_N(U)\right| \ge \frac{\mathbb{E}M_N(U)}{2}\right]\\ &\le 4e^{-2N\delta} \le e^{-N\delta},\end{aligned}$$

proving the claim.

The estimate (3.11) is evident from Sanov's theorem, for any neighborhood U of ν, so it remains to prove (3.12). We claim that

$$\lim_{r\to 0}\inf_{\rho,\theta\in \mathrm{cl}B_r(\nu):\rho^{(j)}=\theta^{(j)}}\left\{H(\rho \mid \mu) + H\left(\theta \mid \theta^{(j)}\otimes\hat{\mu}^{(j)}\right)\right\} = H(\nu \mid \mu) + H\left(\nu \mid \nu^{(j)}\otimes\hat{\mu}^{(j)}\right), \tag{3.13}$$

where $\hat{\mu}^{(j)} \stackrel{\mathrm{def}}{=} \mu_{j+1}\otimes\cdots\otimes\mu_n$. The inequality $\le$ is evident, and the opposite follows from the semicontinuity properties: One gets that for a sequence (ρ_n, θ_n) with $\rho_n^{(j)} = \theta_n^{(j)}$ and $\rho_n, \theta_n \to \nu$, we have

$$\begin{aligned}\liminf_{n\to\infty} H(\rho_n \mid \mu) &\ge H(\nu \mid \mu),\\ \liminf_{n\to\infty} H\left(\theta_n \mid \theta_n^{(j)}\otimes\hat{\mu}^{(j)}\right) &\ge H\left(\nu \mid \nu^{(j)}\otimes\hat{\mu}^{(j)}\right),\end{aligned}$$

the first inequality by the standard semi-continuity, and the second by Lemma 3.6. This proves (3.13).

Choose $\eta > 0$ such that $H\left(\nu^{(j)} \mid \mu^{(j)}\right) < \Gamma_j \log 2 - \eta$, for all $1 \le j \le n$. By (3.13) we may choose r small enough such that for all $1 \le j \le n$,

$$\begin{aligned}\inf_{\rho,\theta \in \mathrm{cl}B_r(\nu):\rho^{(j)}=\theta^{(j)}} & \left\{H(\rho \mid \mu) + H\left(\theta \mid \theta^{(j)} \otimes \hat{\mu}^{(j)}\right)\right\} \\ & \ge H(\nu \mid \mu) + H\left(\nu \mid \nu^{(j)} \otimes \hat{\mu}^{(j)}\right) - \eta/2 \\ & = 2H(\nu \mid \mu) - H\left(\nu^{(j)} \mid \mu^{(j)}\right) - \eta/2 \\ & \ge 2H(\nu \mid \mu) - \Gamma_j \log 2 + \eta/2.\end{aligned}$$

Recall that for two indices α, α' we write $q(\alpha, \alpha') = \max\left\{j : \alpha^{(j)} = \alpha'^{(j)}\right\}$ and $\max \emptyset = 0$. Then

$$\begin{aligned}\mathbb{E}M_N^2(U) &= \sum_{j=0}^{n} \sum_{\alpha,\alpha':q(\alpha,\alpha')=j} \mathbb{P}\left[L_{N,\alpha} \in U, L_{N,\alpha'} \in U\right] \\ &= \sum_{\alpha,\alpha':q(\alpha,\alpha')=0} \mathbb{P}\left[L_{N,\alpha} \in U\right] \mathbb{P}\left[L_{N,\alpha'} \in U\right] + \\ &\quad + \sum_{j=1}^{n} \sum_{\alpha,\alpha':q(\alpha,\alpha')=j} \mathbb{P}\left[L_{N,\alpha} \in U, L_{N,\alpha'} \in U\right] \\ &\le \mathbb{E}[M_N(\mathrm{cl}U)]^2 + \\ &\quad + \sum_{j=1}^{n} \sum_{\alpha,\alpha':q(\alpha,\alpha')=j} \mathbb{P}\left[L_{N,\alpha} \in \mathrm{cl}U, L_{N,\alpha'} \in \mathrm{cl}U\right].\end{aligned} \tag{3.14}$$

We write the empirical measure in the form (3.10), and can use Lemma 3.7. For any $1 \le j \le n$ we have

$$\begin{aligned}\sum_{\alpha,\alpha':q(\alpha,\alpha')=j} & \mathbb{P}\left[L_{N,\alpha} \in \mathrm{cl}U, L_{N,\alpha'} \in \mathrm{cl}U\right] = \\ & = 2^{\Gamma_j N} 2^{(1-\Gamma_j)N} \left(2^{(1-\Gamma_j)N} - 1\right) \mathbb{P}\left[L_{N,\alpha} \in \mathrm{cl}U, L_{N,\alpha'} \in \mathrm{cl}U\right],\end{aligned}$$

where on the r.h.s. α, α' is an arbitrary pair with $q(\alpha, \alpha') = j$. Using Lemma 3.7 we have

$$\begin{aligned}&\mathbb{P}\left[L_{N,\alpha} \in \mathrm{cl}U,\ L_{N,\alpha} \in \mathrm{cl}U\right] \\ &\le \exp\left[-N \inf_{\rho,\theta \in \mathrm{cl}U,\rho^{(j)}=\theta^{(j)}} \left\{H\left(\rho^{(j)} \mid \mu^{(j)}\right) + H\left(\rho \mid \rho^{(j)} \otimes \hat{\mu}^{(j)}\right) + H\left(\theta \mid \theta^{(j)} \otimes \hat{\mu}^{(j)}\right)\right\} + \frac{N\eta}{4}\right] \\ &= \exp\left[-N \inf_{\rho,\theta \in \mathrm{cl}U,\rho^{(j)}=\theta^{(j)}} \left\{H(\rho \mid \mu) + H\left(\theta \mid \theta^{(j)} \otimes \hat{\mu}^{(j)}\right)\right\} + \frac{N\eta}{4}\right] \\ &\le 2^{\Gamma_j N} \exp\left[-2NH(\nu \mid \mu) - \frac{N\eta}{4}\right],\end{aligned}$$

and thus

$$\sum_{\alpha,\alpha':q(\alpha,\alpha')=j} \mathbb{P}\left[L_{N,\alpha} \in \mathrm{cl}U,\ L_{N,\alpha} \in \mathrm{cl}U\right] \le 2^{2N} \exp\left[-2NH(\nu \mid \mu) - \frac{N\eta}{4}\right].$$

Combining, we obtain

$$\operatorname{var}[M_N(U)] \leq 2^{2N} \exp\left[-2NH(\nu \mid \mu) - \frac{N\eta}{4}\right] \leq e^{-N\eta/4}\mathbb{E}[M_N(U)]^2,$$

which proves our claim. $\square$

Proof of Theorem 3.1. We set

$$\mathcal{G} \stackrel{\text{def}}{=} \left\{\nu \in \mathcal{M}_1^+(S^n) : H\left(\nu^{(j)} \mid \mu^{(j)}\right) \leq \Gamma_j \log 2, j = 1, \ldots, n\right\},$$

which is evidently a compact set.

I) We first prove the lower bound. By compactness of $\mathcal{G}$ and the semicontinuity of H there exists $\nu_0 \in \mathcal{G}$ such that

$$\sup_{\nu\in\mathcal{G}} \{\Phi(\nu) - H(\nu \mid \mu)\} = \Phi(\nu_0) - H(\nu_0 \mid \mu).$$

We set $\nu_\lambda \stackrel{\text{def}}{=} (1-\lambda)\nu_0 + \lambda\mu$ for $0 < \lambda < 1$. By convexity of $H(\nu \mid \mu)$ in ν we see that $H\left(\nu_\lambda^{(j)} \mid \mu^{(j)}\right) < H\left(\nu_0^{(j)} \mid \mu^{(j)}\right)$ for all $1 \leq j \leq n$. Furthermore $\nu_\lambda \to \nu_0$ weakly as $\lambda \to 0$, and $\Phi(\nu_\lambda) \to \Phi(\nu_0)$, $H(\nu_\lambda \mid \mu) \to H(\nu_0 \mid \mu)$.
Given $\varepsilon > 0$ we choose $\lambda > 0$ such that

$$\Phi(\nu_\lambda) - H(\nu_\lambda \mid \mu) \geq \Phi(\nu_0) - H(\nu_0 \mid \mu) - \varepsilon.$$

By the continuity of Φ and Lemma 3.10 we find a neighborhood U of ν_λ, and $\delta > 0$ such that

$$\left|\Phi(\theta) - \Phi(\nu_\lambda)\right| \leq \varepsilon, \theta \in U,$$

and

$$\mathbb{P}\left[M_N(U) \leq 2^N \exp\left[-NH(\nu_\lambda \mid \mu) - N\varepsilon\right]\right] \leq e^{-\delta N},$$

for large enough N. Then, with probability greater than $1 - e^{-\delta N}$,

$$\begin{aligned} Z_N &= 2^{-N}\sum_\alpha \exp\left[N\Phi(L_{N,\alpha})\right] \\ &\geq \exp\left[N\Phi(\nu_\lambda) - N\varepsilon\right]\exp\left[-NH(\nu_\lambda \mid \mu) - N\varepsilon\right] \\ &\geq \exp\left[N\sup_{\nu\in\mathcal{G}}\{\Phi(\nu) - H(\nu \mid \mu)\} - 3N\varepsilon\right]. \end{aligned}$$

By Borel-Cantelli, we therefore get, as ε is arbitrary,

$$\liminf_{N\to\infty} \frac{1}{N}\log Z_N \geq \sup_{\nu\in\mathcal{G}}\{\Phi(\nu) - H(\nu \mid \mu)\}$$

almost surely.

II) We prove the upper bound. Let again $\varepsilon > 0$ and set

$$\overline{\mathcal{G}} \stackrel{\text{def}}{=} \{\nu : H(\nu \mid \mu) \leq \log 2\}.$$

If $\nu \in \mathcal{G}$ we choose $r_\nu > 0$ such that $|\Phi(\theta) - \Phi(\nu)| \leq \varepsilon$, $\theta \in B_{r_\nu}(\nu)$ and

$$\mathbb{P}\left[M_N(B_{r_\nu}(\nu)) \geq 2^N \exp\left[-NH(\nu \mid \mu) + N\varepsilon\right]\right] \leq e^{-N\delta_\nu},$$

for some $\delta_\nu > 0$ and large enough N (using Lemma 3.8). If $\nu \in \overline{\mathcal{G}} \setminus \mathcal{G}$ we choose r_ν such that $|\Phi(\theta) - \Phi(\nu)| \leq \varepsilon$, $\theta \in B_{r_\nu}(\nu)$, and

$$\mathbb{P}\left[M_N(B_{r_\nu}(\nu)) \neq 0\right] \leq e^{-N\delta_\nu}, \tag{3.15}$$

again for large enough N (and by Lemma 3.9). As $\overline{\mathcal{G}}$ is compact, we can cover it by a finite union of such balls, i.e.

$$\overline{\mathcal{G}} \subset U \stackrel{\text{def}}{=} \bigcup_{l=1}^{m} B_{r_j}(\nu_j),$$

where $r_j \stackrel{\text{def}}{=} r_{\nu_j}$. We also set $\delta \stackrel{\text{def}}{=} \min_j \delta_{\nu_j}$. We then estimate

$$Z_N \leq 2^{-N} \sum_{l=1}^{m} \sum_{\alpha: L_{N,\alpha} \in B_{r_l}(\nu_l)} \exp\left[N\Phi(L_{N,\alpha})\right] + 2^{-N} \sum_{\alpha: L_{N,\alpha} \notin U} \exp\left[N\Phi(L_{N,\alpha})\right]. \tag{3.16}$$

we first claim that almost surely the second summand vanishes provided N is large enough, i.e. that there is no α with $L_{N,\alpha} \notin U$. By Sanov's theorem, we have

$$\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}\left[L_{N,\alpha} \notin U\right] \leq -\inf_{\nu \notin U} H(\nu \mid \mu) < -\log 2.$$

Therefore, almost surely, there is no α with $L_{N,\alpha} \notin U$, and therefore the second summand in (3.16) vanishes for large enough N, almost surely. The same applies to those summands in the first part of (3.16) for which $\nu_l \notin \mathcal{G}$, using (3.15). We therefore have, almost surely, for large enough N,

$$\begin{aligned} Z_N &\leq 2^{-N} \sum_{l:\nu_l \in \mathcal{G}} \sum_{\alpha: L_{N,\alpha} \in B_{r_l}(\nu_l)} \exp\left[N\Phi(L_{N,\alpha})\right] \\ &\leq e^{N\varepsilon} \sum_{l:\nu_l \in \mathcal{G}} \exp\left[N\Phi(\nu_l)\right] M_N(B_{r_l}(\nu_l)) \\ &\leq e^{2N\varepsilon} \sum_{l:\nu_l \in \mathcal{G}} \exp\left[N\Phi(\nu_l)\right] \exp\left[-NH(\nu_l \mid \mu)\right] \\ &\leq e^{2N\varepsilon} m \exp\left[N \sup_{\nu \in \mathcal{G}} \{\Phi(\nu) - H(\nu \mid \mu)\}\right]. \end{aligned}$$

As ε is arbitrary, we get

$$\limsup_{N\to\infty} \frac{1}{N} \log Z_N \leq \sup_{\nu \in \mathcal{G}} \big\{\Phi(\nu) - H(\nu \mid \mu)\big\}.$$

This finishes the proof of Theorem 3.1. □

3.4.1. Gibbs vs. Parisi. To goal of this section is to prove the equivalence of the Parisi Variational Principle, PVP for short, and the GVP. We sketch the strategy.

Recall that the PVP is the minimization problem given by

$$\min_{\boldsymbol{m}: 0 \leq m_1 \leq \cdots \leq m_n \leq 1} \mathsf{Parisi}(\phi, \boldsymbol{m}, \mu). \tag{3.17}$$

By Hölder inequality it is easily seen that the map

$$\mathbb{R}^n \ni \boldsymbol{t} \mapsto \mathsf{Parisi}\left(\phi, \frac{1}{t_1}, \frac{1}{t_2}, \dots, \frac{1}{t_n}, \mu\right)$$

is convex, a simple observation which entails uniqueness of the minimizer of the PVP. A somewhat less evident observation concerns the partial derivatives (in the variables $m_1, \dots, m_n$) of the functional $\mathsf{Parisi}(\phi, \boldsymbol{m}, \mu)$, which turn out to be related to the relative entropies associated to the generalized Gibbs measure $G(\phi, \boldsymbol{m}, \mu)$, cfr. Proposition 3.11 below. Given this interpretation, it steadily follows that the Generalized Gibbs measure associated to the minimizer $\underline{\boldsymbol{m}}$ satisfies the side constraints in the GVP. We then step back to the GVP, claiming that any measure $\nu \in \mathcal{M}_1^+(S^n)$ satisfying the side constraints of the GVP and outbeating the Gibbsian $G(\phi, \mu, \underline{\boldsymbol{m}})$ must have negative relative entropy, $H(\nu \mid G(\phi, \mu, \underline{\boldsymbol{m}})) \leq 0$, cfr. Proposition 3.12. By positivity of relative entropies, Theorem 3.3 then follows.

We begin providing the link between the GVP and PVP. Denote by $G = G(\phi, \boldsymbol{m})$ the generalized Gibbs measure associated to $\boldsymbol{m} \in \mathbb{R}^n$, by $G^{(j)}$ its marginal on the first j coordinates and by $\check{G}^{(j)}$ the conditional distribution on the j-coordinate given the first $j-1$.

Proposition 3.11. *For $j = 1, \dots, n$ it holds:*

$$\partial_{m_j} \mathsf{Parisi}(\phi, \mathbf{m}, \mu) = \frac{1}{m_j^2}\left\{\mathsf{E}_{G^{(j-1)}}\left[H\big(\check{G}^{(j)} \mid \mu_j\big)\right] - \gamma_j \log 2\right\}$$

Proof. Follows plainly from the definition of generalized Gibbs measure. □

The second ingredient to establish the equivalence between GVP and PVP is the following observation on relative entropies: for $\nu \in \mathcal{M}_1^+(S^n), \boldsymbol{m} \in \mathbb{R}^n$ and generalized Gibbs measure $G = G(\phi, \boldsymbol{m}, \mu)$, we set

$$\Delta(G, \nu) \stackrel{\text{def}}{=} \mathsf{Gibbs}(\phi, G, \mu) - \mathsf{Gibbs}(\phi, \nu, \mu), \Delta H_i \stackrel{\text{def}}{=} H(G^{(i)} \mid \mu^{(i)}) - H(\nu^{(i)} \mid \mu^{(i)}),$$

and finally $\chi_j = \chi_j(\boldsymbol{m}) \stackrel{\text{def}}{=} \dfrac{m_j - m_{j+1}}{m_j} \cdot \dfrac{m_n}{m_{j+1}}$.

Proposition 3.12. *With the above notations,*

$$H(\nu \mid G) = m_n \Delta(G, \nu) + (m_n - 1)\Big\{H(G \mid \mu) - H(\nu \mid \mu)\Big\} + \\ + \sum_{j=1}^{n-1} \chi_j \Delta H_j + \sum_{i=1}^{n} (m_i - m_{i+1}) \sum_{j=1}^{i} \frac{1}{m_j} \mathsf{E}_{\nu^{(j-1)}}\left[H\left(\check{\nu}^{(j)} \middle| \check{G}^{(j)}\right)\right].$$

We postpone the proof of Proposition 3.12, and show first how to combine it with Propositions 3.11 to obtain the equivalence of GVP and PVP, and we start with the latter: the target function $\mathsf{Parisi}(\phi, \cdot)$ to minimize on the compact set $\{\boldsymbol{m} \in \mathbb{R}^n : 0 \leq m_1 \leq m_2 \cdots \leq m_n \leq 1\}$ is continous everywhere except on the "0-boundary", i.e. when $m_1 = m_2 = \cdots = m_j = 0$ for some $j = 1, \dots, n$, in which case it runs off to $+\infty$. Moreover, an easy application of Hölder's inequality shows that the function

$\boldsymbol{t} \in \mathbb{R}^n \mapsto \mathsf{Parisi}(\phi, \frac{1}{t_1}, \frac{1}{t_2}, \ldots, \frac{1}{t_n})$ is convex, entailing uniqueness of the minimizer of the PVP,

$$\underline{\boldsymbol{m}} \stackrel{\text{def}}{=} \arg \inf_{0 \le m_1 \le \cdots \le m_n} \left\{ \sum_{i=1}^n \gamma_i \frac{\log 2}{m_i} + \mathcal{P}\left(\phi, \mu, m_1, \ldots, m_n\right) - \log 2 \right\}.$$

By Real Analysis considerations, a necessary condition for minimality is that there are $K \in \{1, \ldots, n\}$, a strictly increasing sequence of reals $0 < m^{(1)} < m^{(2)} < \cdots < m^{(K)} < m^{(K+1)} = 1$, and integers $0 = j_0 < j_1 < j_2 < \cdots < j_K < j_{K+1} = n$ such that the choice $\underline{m_l} \equiv m^{(r)}$ for $l = j_{r-1}+1, \ldots, j_r$ and $r = 1, \ldots, K+1$ is optimal (minimal). By standard real analysis considerations we then have that for any integer $r = 1, \ldots, K$ the following directional derivative (inside of the complete "block") vanishes:

$$\frac{d}{dt}\Big|_{t=0} \mathsf{Parisi}\big(\phi, m_1, m_2, \ldots, m_{j_{r-1}+1} + t, \ldots, m_{j_r} + t, m_{j_r+1}, \ldots, \ldots, m_n\big)\Big|_{\boldsymbol{m} = \underline{\boldsymbol{m}}} = 0, \tag{3.18}$$

which by Proposition 3.11 is equivalent to

$$\sum_{k=j_{r-1}+1}^{j_r} \mathsf{E}_{G^{(k-1)}} \Big[H\big(\check{G}^{(k)} \mid \mu_k\big)\Big] = \gamma^{(r)} \log 2, \tag{3.19}$$

where $\gamma^{(r)} \stackrel{\text{def}}{=} \sum_{i=j_{r-1}}^{j_r} \gamma_i$. On the other hand, for $l < j_r - j_{r-1}$ strictly and $r = 1, \ldots, K+1$ we have

$$\frac{d}{dt}\Big|_{t=0} \mathsf{Parisi}\big(\phi, m_1 + t, m_2 + t, \ldots, m_{j_{r-1}+l} + t, m_{j_{r-1}+l+1}, \ldots, m_{j_r}, \ldots, m_n\big)\Big|_{\boldsymbol{m} = \underline{\boldsymbol{m}}} \le 0,$$

which by Proposition 3.11 and (3.18) implies that

$$\sum_{k=j_{r-1}+1}^{j_{r-1}+l} \mathsf{E}_{G^{(k-1)}} \Big[H\big(\check{G}^{(k)} \mid \mu_k\big)\Big] \le \Gamma_l \log 2. \tag{3.20}$$

Combining (3.19), (3.20) we see that the generalized Gibbs measure $G = G(\phi, \underline{\boldsymbol{m}})$ is such that

$$\begin{aligned} &i)\ H(G^{(j)} \mid \mu^{(j)}) \le \Gamma_j \log 2 \quad \text{for all } j = 1, \ldots, n; \\ &ii)\ H(G^{(j_r)} \mid \mu^{(j_r)}) = \Gamma_{j_r} \log 2 \quad \text{for } r = 1, \ldots, K. \end{aligned} \tag{3.21}$$

Satisfying the constraints in the GVP, such a generalized Gibbs measure clearly yields a lower bound.

We now claim that, in fact, it is optimal. To see this, suppose to have $\nu \in \mathcal{M}_1^+(S^n)$ satisfying the side conditions of the GVP and "outbeating" the Gibbs measure $G(\phi, \underline{\boldsymbol{m}})$ in the sense that $\mathsf{Gibbs}(\phi, \nu, \mu) \ge \mathsf{Gibbs}(\phi, G(\phi, \underline{\boldsymbol{m}}, \mu), \mu)$, and thus $\Delta(G, \nu) \le 0$. By Proposition 3.12,

$$\begin{aligned} H(\nu \mid G(\phi, \underline{\boldsymbol{m}})) &= \underline{m}_n \Delta(G, \nu) + (\underline{m}_n - 1)\Big\{H(G|\mu) - H(\nu|\mu)\Big\} + \\ &+ \sum_{j=1}^{n-1} \chi_j(\underline{\boldsymbol{m}}) \Delta H_j + \sum_{i=1}^n (\underline{m}_i - \underline{m}_{i+1}) \sum_{j=1}^i \frac{1}{\underline{m}_j} \mathsf{E}_{\nu^{(j-1)}} \left[H\left(\check{\nu}^{(j)} \Big| \check{G}^{(j)}\right)\right]. \end{aligned} \tag{3.22}$$

For $j, j+1$ both inside a given box (i.e. in the set $\{j_{r-1}+1, \dots, j_r\}$ for some $r = 1, \dots, K+1$), we always have $\underline{m}_j - \underline{m}_{j+1} = 0$. So, the last term on the r.h.s of (3.22) reads

$$\sum_{r=1}^{K+1} \left(m^{(r)} - m^{(r+1)}\right) \sum_{j=1}^{j_r} \frac{1}{\underline{m}_j} \mathsf{E}_{\nu^{(j-1)}} \left[H\left(\check{\nu}^{(j)} \middle| \check{G}^{(j)}\right) \right],$$

and thus ≤ 0 by positivity of relative entropies and the fact that $m^{(j)}$ is strictly increasing. On the other hand, the function χ's are such that $\chi_j(\underline{\boldsymbol{m}}) = 0$ if $j, j+1$ both fall in a box and ≤ 0 if $j = j_r$ for some $r = 1, \dots, K$, so the contribution of the term before last on the r.h.s of (3.22) is given by

$$\sum_{r=1}^{K} \chi_{j_r} \Delta H_{j_r} \overset{\text{ii) of (3.21)}}{=} \sum_{r=1}^{K} \chi_{j_r} \left(\frac{j_r}{n} \log 2 - H\left(\nu^{(j_r)} \mid \mu^{(j_r)}\right) \right) \leq 0,$$

since the measure ν must satisfy the constraints and in particular $H\left(\nu^{(j_r)} \mid \mu^{(j_r)}\right) \leq \frac{j_r}{n} \log 2$, for every $r = 1, \dots, K+1$. The second term on the r.h.s of (3.22) is also negative: either is $\underline{m}_n = 1$, in which case it simply vanishes, or $\underline{m}_n < 1$, in which case it reads

$$(\underline{m}_n - 1)\Big[\log 2 - H(\nu \mid \mu)\Big] \leq 0,$$

since on the one hand, by *ii)* of (3.21) we have that $H(G \mid \mu) = \log 2$, and on the other hand the measure ν must satisfy the side constraint $H(\nu \mid \mu) \leq \log 2$. As for the first term on the r.h.s of (3.22), it is negative by assumption that the measure ν outbeats the generalized Gibbs measure $G(\phi, \underline{\boldsymbol{m}})$. So, every term on the r.h.s of (3.22) is negative: a contradiction. This proves that the generalized Gibbs measure $G(\phi, \underline{\boldsymbol{m}})$ is the unique, extremal measure solving the GVP. It is a simple exercise then to show that $\mathsf{Gibbs}(\phi, G(\phi, \underline{\boldsymbol{m}}), \mu) = \mathsf{Parisi}(\phi, \underline{\boldsymbol{m}}, \mu)$, thus "Gibbs $\equiv$ Parisi".

Proof of Proposition 3.12. Recalling the recursive procedure which to $\phi : S^n \to \mathbb{R}$ and $\mu \in \mathcal{M}_1^+(S^n)$ allows to construct the Generalized Gibbs measure for given vector

$\boldsymbol{m} \in \mathbb{R}^n$, we have:

$$
\begin{aligned}
H(\nu \mid G) &= H(\nu \mid \mu) - \mathsf{E}_\nu\big[\log(dG/d\mu)\big] = \\
&= H(\nu \mid \mu) - \sum_{i=1}^{n} m_i \mathsf{E}_\nu[\phi_i] + \sum_{i=1}^{n} m_i \mathsf{E}_\nu[\phi_{i-1}] = \\
&= (1-m_n)H(\nu \mid \mu) + m_n H(\nu \mid \mu) - m_n \mathsf{E}_\nu[\phi] - \sum_{i=1}^{n-1} m_i \mathsf{E}_\nu[\phi_i] + \sum_{i=1}^{n} m_i \mathsf{E}_\nu[\phi_{i-1}] = \\
&= (1-m_n)H(\nu \mid \mu) - m_n \mathsf{Gibbs}(\phi,\nu,\mu) - \sum_{i=1}^{n-1} m_i \mathsf{E}_\nu[\phi_i] + \sum_{i=1}^{n} m_i \mathsf{E}_\nu[\phi_{i-1}] = \\
&= (1-m_n)H(\nu \mid \mu) + m_n \Delta(G,\nu) - m_n \mathsf{Gibbs}(\phi, G, \mu) + \\
&\qquad\qquad - \sum_{i=1}^{n-1} m_i \mathsf{E}_\nu[\phi_i] + \sum_{i=1}^{n} m_i \mathsf{E}_\nu[\phi_{i-1}] = \\
&= m_n \Delta(G,\nu) + (m_n - 1)\Big\{H(G \mid \mu) - H(\nu \mid \mu)\Big\} + \\
&\qquad - m_n \mathsf{E}_G[\phi] + H(G \mid \mu) - \sum_{i=1}^{n-1} m_i \mathsf{E}_\nu[\phi_i] + \sum_{i=1}^{n} m_i \mathsf{E}_\nu[\phi_{i-1}].
\end{aligned}
\tag{3.23}
$$

By the very definition of the generalized Gibbs measure G we have $H(G \mid \mu) = \mathsf{E}_G[\log(dG/d\mu)] = \sum_{i=1}^n \mathsf{E}_G[m_i\phi_i] - \sum_{i=1}^n m_i \mathsf{E}_G[\phi_{i-1}]$, so that plugging this into (3.23), rearranging the sums (and exploiting the fact that $\mathsf{E}_\nu\phi_0 = \mathsf{E}_G\phi_0 = \phi_0$) we have

$$
\begin{aligned}
H(\nu \mid \mu) &= m_n \Delta(G,\nu) + (m_n - 1)\Big\{H(G \mid \mu) - H(\nu \mid \mu)\Big\} + \\
&\qquad + \sum_{i=1}^{n-1} (m_i - m_{i+1})\Big\{\mathsf{E}_G[\phi_i] - \mathsf{E}_\nu[\phi_i]\Big\}.
\end{aligned}
\tag{3.24}
$$

Remark that

$$
\begin{aligned}
\mathsf{E}_\nu\Big[H(\check\nu^{(i)} \mid \check G^{(i)})\Big] &= \mathsf{E}_\nu\Big[H(\check\nu^{(i)} \mid \mu_i)\Big] - \mathsf{E}_\nu\Big[H(\check G^{(i)} \mid \mu_i)\Big] \\
&= \mathsf{E}_\nu\Big[H(\check\nu^{(i)} \mid \mu_i)\Big] - m_i \mathsf{E}_\nu[\phi_i] + m_i \mathsf{E}_\nu[\phi_{i-1}],
\end{aligned}
\tag{3.25}
$$

and similarly

$$
\mathsf{E}_G[H(\check G^{(i)} \mid \mu_i)] = m_i \mathsf{E}_G[\phi_i] - m_i \mathsf{E}_G[\phi_{i-1}]. \tag{3.26}
$$

Setting $\delta_i \overset{\text{def}}{=} \mathsf{E}_G[\phi_i] - \mathsf{E}_\nu[\phi_i]$, combining (3.25) and (3.26) we thus get the recursion $\delta_i = \delta_{i-1} + \frac{1}{m_i}\Upsilon_i$ for any $i = 1, \ldots, n$, with

$$
\begin{aligned}
\Upsilon_i &\overset{\text{def}}{=} \mathsf{E}_\nu\Big[H(\check\nu^{(i)} \mid \check G^{(i)})\Big] + \mathsf{E}_G[H(\check G^{(i)}) \mid \mu_i)] - \mathsf{E}_\nu\Big[H(\check\nu^{(i)} \mid \mu_i)\Big] = \\
&= \mathsf{E}_\nu\Big[H(\check\nu^{(i)} \mid \check G^{(i)})\Big] + \Big\{\Delta H_i - \Delta H_{i-1}\Big\}.
\end{aligned}
\tag{3.27}
$$

Since $\delta_0 = 0$, we have that $\delta_i = \sum_{j=1}^{i} \frac{1}{m_j} \Upsilon_j$, and plugging this into (3.24) we obtain

$$H(\nu \mid G) = m_n \Delta(G,\nu) + (m_n - 1)\Big\{H(G \mid \mu) - H(\nu \mid \mu)\Big\} + \\ + \sum_{i=1}^{n-1} (m_i - m_{i+1}) \sum_{j=1}^{i} \frac{1}{m_j} \Upsilon_j. \tag{3.28}$$

This, together with the second line in (3.27) yields the claim of Proposition 3.12. □

Proof of Theorem 3.4. Recall that $F_N(\beta) = F_N(\beta, X)$ where X is given by the hamiltonian (3.4). The fact that the free energy is *self-averaging*, meaning that $f(\beta)$ (if the limit exists) is also the almost sure limit of the F_N is a simple consequence of the Gaussian concentration inequality. We write F_N as a function of the standardized variables $\widetilde{X}_{\alpha_1,\dots,\alpha_j} \stackrel{\text{def}}{=} X_{\alpha_1,\dots,\alpha_j}/\sqrt{a_j N}$ and $g_{\alpha_1,\dots,\alpha_j,k}$. As

$$\left|\log \sum\nolimits_i \mathrm{e}^{a_i} - \log \sum\nolimits_i \mathrm{e}^{a_i'}\right| \le \max_i \left|a_i - a_i'\right|, \ a_i, a_i' \in \mathbb{R},$$

we get that $F_N(\beta)$, regarded as a function of the collection $\left(\widetilde{X}_{\alpha_1,\dots,\alpha_j}, g_{\alpha_1,\dots,\alpha_j,k}\right)$ is Lipschitz continuous with Lipschitz constant $\beta/\sqrt{N}$. By the usual concentration of measure estimates for Gaussian distributions, see e.g. [**15**, Proposition 2.18], we have

$$\mathbb{P}\Big[\big|F_N(\beta) - \mathbb{E}F_N(\beta)\big| > \epsilon\Big] \le 2\exp\left[-\frac{\epsilon^2}{2\beta^2} N\right] \tag{3.29}$$

Using the Borel-Cantelli lemma, one sees that if $\lim_{N\to\infty} f_N(\beta)$ exists, then the $F_N(\beta)$ converge almost surely to this limit, too, and if $\lim_{N\to\infty} F_N(\beta)$ exists almost surely, then the limit is non-random and equals $\lim_{N\to\infty} f_N(\beta)$.

As for the computation of the $\mathbb{P}$-a.s. limit, we first carry out the trace over the Ising spins:

$$F_N(\beta) = \frac{1}{N} \log 2^{-N} \sum_{\alpha\in\Sigma_N^{\text{grem}}} \exp\left[\beta X_\alpha + \sum_{i=1}^{N} \log\cosh(\beta g_{\alpha,i})\right]$$

and then express, for $\alpha \in \Sigma_N^{\text{grem}}$, the GREM-part as $X_\alpha = \sum_{i=1}^N X_{\alpha,i}$ for centered gaussians such that $\mathbb{E}[X_{\alpha,i} X_{\alpha',j}] = \delta_{i=j} q(\alpha,\alpha')$. Let

$$\widetilde{F}_N(\beta) \stackrel{\text{def}}{=} \frac{1}{N} \log 2^{-N} \sum_{\alpha\in\Sigma_N^{\text{grem}}} \exp\Big[\sum_{i=1}^{N} \phi\big(X_{\alpha_1,i}, g_{\alpha_1,i}, \dots, X_{\alpha_1,\dots,\alpha_n,i}, g_{\alpha_1,\dots,\alpha_n,i}\big)\Big]$$

with $\phi : \mathbb{R}^{2n} \to \mathbb{R}, \phi\Big((x_1,y_1),\dots,(x_n,y_n)\Big) \stackrel{\text{def}}{=} \sum_{k=1}^n \beta x_k + \log\cosh(\beta \sum_{k=1}^n y_k)$. $F_N(\beta)$ and $\widetilde{F}_N(\beta)$ are equally distributed, thus $\lim_{N\to\infty} F_N(\beta) = \lim_{N\to\infty} \widetilde{F}_N(\beta)$, whenever the second limit exists; this is however the case, since by Theorem 3.1, and the equivalence of GVP and PVP we have that $\widetilde{F}_N(\beta)$ converges $\mathbb{P}$-almost surely to $f(\beta)$ as given by formula (3.6). □

Bibliography

[1] M. Aizenman, R.Sims and S.Starr, *Extended variational principle for the Sherrington-Kirkpatrick spin-glass model*, Phys. Rev. B 68, 214403 (2003).

[2] A.D.Barbour, L.Holst and S. Janson, *Poisson Approximation*, Clarendon Press Oxford (1992).

[3] E.Bolthausen and N.Kistler, *On a non hierarchical version of the Generalized Random Energy model*, Ann. Appl. Probab. 16, no. 1 (2006).

[4] E. Bolthausen and A.S. Sznitman, *Ten lectures on Random Media*, DMV Seminar, Birkhäuser (2001).

[5] E. Bolthausen and A.S. Sznitman, *On Ruelle's probability cascades and an abstract cavity method*, Comm. Math. Phys. 197 (1998).

[6] A. Bovier and I. Kurkova *Derrida's Generalized Random Energy Models I & II.* Annals de l'Institut Henri Poincaré **40**, 439-495 (2004).

[7] M. Bramson, *Maximal displacement of branching Brownian motion*, Comm. Pure Appl. Math. 31 (1978).

[8] D. Capocaccia, M. Cassandro and P. Picco, *On the existence of thermodynamics for the Generalized Random Energy Model.* J. Stat. Phys. **46,** 493-505 (1987).

[9] M. Cranston and S. Molchanov, *Limit laws for sums of exponentials of iid random variables*, Israel Journal of Math., Vol 148 (2005).

[10] B. Derrida, *Random Energy Model: An exactly solvable model of disordered systems*, Phys. Rev. B 24 (1981).

[11] B. Derrida, *A generalization of the random energy model that includes correlations between the energies*, J.Phys.Lett. **46**, 401-407 (1985).

[12] M.Mézard, G.Parisi and M. Virasoro, *Spin Glass theory and beyond*, World scientific, Singapore (1987).

[13] F. Guerra, *Broken replica symmetry bounds in the mean field spin glass model*, Comm. Math. Phys. 233, 1-12 (2003).

[14] F. Guerra and F.L. Toninelli, *The thermodynamic limit in mean field spin glass models*, Comm. Math. Phys. 230, 71-79 (2002).

[15] M. Ledoux, *The Concentration of Measure Phenomenon*, Mathematical Surveys and Monographs 89, AMS 2002.

[16] M.Mézard, G.Parisi, M. Virasoro, *Spin Glass theory and beyond*, World scientific, Singapore (1987).

[17] V. V. Petrov, *Limit Theorems of Probability Theory*, Oxford Studies in Probability, Clarendon Press (1995).

[18] D. Ruelle, *A mathematical reformulation of Derrida's REM and GREM*, Comm. Math.Phys. 108 (1987).

[19] M.Talagrand, *Spin Glasses: A Challenge for Mathematicians. Cavity and Mean Field Models*, Springer Verlag (2003).

[20] M. Talagrand, *The Parisi formula*, Ann. of Math. (2) 163 (2006).

Curriculum Vitae

Kistler Nicola

Geboren am 17. Juli 1975 in Lugano
Bürger von Reichenburg (SZ)

Ausbildung

Juni 1994: Matura Typus C, Liceo Lugano Uno
WS 1994-SS 1999: Studium der Mathematik an der ETH Zürich
WS 1999/2000: Diplom in Mathematik
Diplomarbeit: "On Generalized Itô's Formulae"
Leitung: Prof.Dr. Francesco Russo (Université Paris 13)

Dissertation

Titel: "On the Ultrametricity in Spin Glass Theory"
Leitung: Prof.Dr. Erwin Bolthausen

Gegenwärtige Stellung

Seit März 2001 Doktorand und Assistent am Institut für Mathematik der Universität Zürich.